A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Consciousness as a Regulatory Field: A Theory of Psychotherapy 275
Rae Madison Collier

The Effects of Fear Arousal and Suppression of Fear Upon Social Perception 283
Seymour Feshbach and Robert D. Singer

The Masculinity-Femininity Dimensions in Normal and Psychotic Subjects ... 289
Max R. Reed

Galvanic Skin Response as a Measure of Patient's Reaction to Therapist's Permissiveness 295
James E. Dittes

The Influence of the Mother's Presence on Children's Doll Play Aggression ... 304
Harry Levin and Valerie F. Turgeon

Verbal Conditioning as a Function of Experimenter Characteristics 309
Arnold Binder, David McConnell, and Nancy A. Sjoholm

Continued on inside front cover

This is the Last Issue of Volume 55
Title Page and Contents for the Volume
Appear Herein.

Volume 55 NOVEMBER 1957 Number 3

CONTENTS—*Continued*

Mediated Generalization and the Incubation Effect as a Function of Manifest Anxiety 315
Martha T. Mednick

A Scale for Measuring Attitudes of Social Responsibility in Children 322
Dale B. Harris

A Cross-Cultural Survey of Some Sex Differences in Socialization 327
Herbert Barry III, Margaret K. Bacon, and Irvin L. Child

Avoidance Learning of Perceptual Defense and Vigilance 333
Don E. Dulany, Jr.

The Relative Aversiveness of Warning Signal and Shock in an Avoidance Situation 339
Murray Sidman and John J. Boren

Selective Recall in Schizophrenia and Its Relation to Ego Strength 345
Arthur S. Tomkins

Effects of Perceived Dependency Relationships Upon Conformity to Group Expectations 350
Leonard Berkowitz

The Effect of Training Upon Accuracy and Variability in Interpersonal Perception 355
W. J. Cross

Reference Groups, Membership Groups, and Attitude Change 360
Alberta Engvall Siegel and Sidney Siegel

Socioeconomic Status and Conceptual Thinking 365
Jerome Siller

Recall of Interrupted Tasks under Stress: A Phenomenon of Memory or of Learning? 372
Albert J. Caron and Michael A. Wallach

Group Climate as a Factor in the Relevance and Accuracy of Social Perception 382
Ralph V. Exline

CRITIQUE AND NOTES

On the Phenomena of "Opposite Speech" 389
Bernard Kaplan

Ethnocentrism and Stimulus Generalization: A Replication and Further Study 393
Franklyn N. Arnoff

Social Desirability and Probability of Endorsement of Items in the Interpersonal Check List 394
Allen L. Edwards

Social Setting and Conformity to a Legal Requirement 396
Abram M. Barch, Don Tresise, and John Naugle

Public and Private Conformity Under Different Conditions of Acceptance in the Group 398
Herbert Mausd

Resolution of Approach-Approach and Avoidance-Avoidance Conflicts 402
Abe Arkoff

Personality Change as a Result of an Experimental Change of Environmental Conditions 404
Arnold S. Tannenbaum

Volume 55 NOVEMBER, 1957 Number 3

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

GARDNER LINDZEY
University of Minnesota

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: BARBARA CUMMINGS, FRANCES H. CLARK, SADIE J. DOYLE, SARAH WOMACK

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Change of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1957 by the American Psychological Association, Inc.

CONSCIOUSNESS AS A REGULATORY FIELD: A THEORY OF PSYCHOTHERAPY

REX MADISON COLLIER

VA Hospital, Jefferson Barracks, Missouri

TWO PREVIOUS papers have outlined and partially developed a regulatory theory of consciousness. The first (3) was a brief initial statement which applied only to nonstress conditions; the second (4) developed the theory in terms of stress-defense dynamics and made applications to basic problems in psychopathology. The present paper proposes to interpret selected aspects of the process of psychotherapy in terms of these concepts.

Any adequate review of the literature on psychotherapy (e.g., Ellis [8]) reveals the growing tendency to eclecticism. In any field, eclecticism implies the need for a set of basic concepts that bring system, relatedness, and even unity to the common factors empirically selected from a variety of sources. Among the few current attempts to integrate the varied phenomena of psychotherapy in terms of a systematic set of concepts, the most notable have originated from Hullian learning theory (7, 12, 14, 15, 16). The chief difference between this type of theory and the kind here being developed is probably one of emphasis. Hullian theory emphasizes peripheral factors for the sake of accessibility and objectivity. The present theory places more emphasis on central factors with utilization of the biological concept of regulation. Some reduction in objectivity is risked for the purpose of attempting an organized approach which gives a more appropriate place to the unsuccessfully rejected concept of consciousness.

Any theory of how psychotherapy functions must be based on concepts of integration or the broader concept of regulation. The problem of psychotherapy may be stated as that of discovering and managing the conditions which facilitate the re-establishment of a more adequate degree of self-regulation. It is assumed that no definition or approach to psychotherapy is complete or sufficiently basic without including the concept of self-regulation in some form. The assumptions (*a*) that consciousness has a functional role in behavior and (*b*) that the concepts of field and regulation are essential to this function not only give new meaning to the neglected but inescapable concept of consciousness but also lay the basis for developing and defining an interrelated group of dynamic concepts that seem to be basic in behavior.

While some writers (e.g., John Dewey [6]) have denied that consciousness can be successfully defined, for the purposes of this theory the following definition is proposed: Consciousness is an awareness field of biopsychological origin in which processes which may eventuate in action have increased chances of interaction and mutual modification. Thus, through such an awareness field, however simple and rudimentary or complex and comprehensive, the organism is provided with the means for a more flexible adaptation. In this sense, therefore, consciousness assumes a role which gives it great survival value in the evolutionary process. Since the problem of psychotherapy is to a large extent one of reorganization and readjustment, it is no mere coincidence that progress in reorganization is closely related to the degree to which the individual can allow himself to become aware of significant issues.

The regulatory effectiveness of the conscious field is dependent upon its magnitude which varies with the degree to which processes relevant to the current issues may be accessible to the field. A reduction of this kind of magnitude through blocking, resistance, or repression results in some degree of coincident reduction in the flexibility and regulatory adequacy of the conscious field. It is assumed that the more permeable the conscious field boundaries are to processes relevant to a current issue, the broader and more inclusive is the regulatory frame of reference.

The conditions producing variations in the regulatory effectiveness of the conscious field have been described in terms of five dynamically interrelated conceptualized variables. These are as follows: (*a*) stress, (*b*) defense system with primary and secondary components, (*c*) stress-tolerance thresholds, (*d*) ego strength, and (*e*) boundary permeability of the conscious field. These concepts, defined and

utilized in detail in a previous paper (4) lead to a central concept of stress-defense dynamics. In brief schematic form, the problem of psychotherapy in terms of the regulatory theory of consciousness and these related variables may be stated as follows:

1. The psychopathology which the individual (patient[1] or client) seeks to reduce is a product of the individual's system of defense interacting with experienced stress.

2. As a result of the stress-defense interaction, the regulatory field of consciousness becomes restricted to a degree that cripples its regulatory effectiveness. This allows the rejected segments of experience and of the self to produce biased and poorly managed impulses and behavior which often dominate responses to the further discomfiture of both the individual and his associates.

3. Assuming that the pathological processes are reversible to some degree, the first or primary essential in psychotherapy is the provision of a relationship designed to reduce the individual's need for his defensive system and coincidentally to increase his threshold of tolerance for stress. Therefore, some degree of dependence of the individual on the therapist for some period of time is necessary in most instances.

4. Communication between the individual and the therapist facilitates intrapersonal communication, which in turn increases chances for greater permeability of the restricting defensive boundaries to the conscious field.

5. Increases in ego strength—the ability to accept into consciousness potentially stressful material and to deal with it constructively—occur as the individual experiences acceptance by another who is significant to him. Thus, the individual learns to accept himself and thereby lays the basis for a more adequate degree of independence and self-regulation.

6. The rate at which rejected segments of experience and self can be returned to the regulatory conscious field depends upon all of these interrelated factors taken together, but initially the most important factor is the therapeutic relationship.

7. Increases in ego strength are of great importance, for they are associated with the following changes: (*a*) higher stress-tolerance thresholds, (*b*) reduction of need for the defense system, (*c*) increased permeability of the boundaries to the conscious field, and (*d*) increased self understanding or insight. Taken together, these results tend toward greater emotional and intellectual freedoms to move in directions more representative of the total personality.

It is assumed that there is no final or essential conflict between this theoretical approach and the established facts of learning. Some learning theories will emphasize particulars and their constancies, e.g., S-R, cues, verbal symbols, etc., while here the emphasis is on central regulation, field interaction, and the mutual modifiability of the interacting processes. Some may worry about the subjectivity aspect of the concept of consciousness, but it is important to note that even some behavioristic theories have much need for the intervening variable. The conscious field may be thought of, if one wishes, as a field of interacting intervening variables with regulatory significance for both the S and the R. This will emphasize, however, the "O" of Woodworth's S–O–R formula (17) and reduce the emphasis on S–R aspects. Often the essential difference between theories is the place of emphasis as one plays the scientific game of attempting to make a subject matter more intelligible.

During any attempt to escape the current eclecticism in psychotherapy and to arrive at a more integrative approach, a set of loosely related, often cumbersome, phenomena will be encountered. A selected group of these to which the regulatory theory of consciousness may be applied is as follows: communication and psychotherapy, free association, insight, effects of "uncovering" and interpretation, abreaction, the relationship aspect of psychotherapy, and group relationship factors. While this is not an exhaustive list, it serves to illustrate how an ordering of the concepts may occur in terms of the regulatory theory of consciousness.

COMMUNICATION AND THERAPY

If the function of the conscious field is regulatory along with maintenance of flexibility of adaptation to immediate issues, then

[1] The term "patient" is used here with special reservations. The traditional connotation that the patient is the passive recipient of things done to him and of orders carried out for him must be rejected. For psychotherapy to be effective, the patient must be an active participant in his own treatment. It is with this connotation that the term is used in this discussion. For discussions of this and related issues reference is made to Lindner (11) and Shakow (13).

the adequacy of this regulation is dependent upon what is available to the field. Cultural predispositions require, for the sake of clarity, a repetition of the point that this concept of consciousness is far broader than a synonym for reasoning, intellectualizing, or problem solving as these terms have been traditionally used. Western culture predisposes the individual to assume that it is "bad taste" to reveal strong feelings. The child who experiences strong emotion arouses anxiety in his elders and soon learns to fear and reject his own feelings. "Control" is thus acquired through an habituated exclusion of affect. The concept of consciousness that tends to be current is in part characterized by "flattened affect." Thus, the group predisposes itself to a culturally based incipient schizophrenia. To whatever extent the individual has denied himself the opportunity or privilege of experiencing his feelings, anxieties, dreads, elations, etc., to that extent he has become a more rigidly organized individual with fewer chances for these feelings to become flexibly related to other areas of his personality. A kind of intrapersonal communication has become interrupted. The function of communication in the process of psychotherapy is to make available to the conscious field, processes relevant to a given issue. Maximal adequacy of self-regulation is associated with maximal intrapersonal communication. Unacceptable experiences and feelings often become rejected and alienated from the main stream of awareness that tends to represent the self. From these rejected segments of experience, feeling, and self-evaluation, there are continued effects in the forms of anxiety, impulsive and compulsive desires, drives, irritations, and feelings of frustration that are poorly integrated with the main trends of personality. These rejected segments of feeling and experience are as rigid and unchanging as is their degree of dissociation from the conscious field. Re-association through improved intrapersonal communication is not only a goal of psychotherapy but is also recognized here as a goal of growth and increased maturity and hence should have an even wider application. One implied application is that the individual should learn to accept and respect the affect aspect of experience even though this may, at times, cause him considerable anxiety.

Since interpersonal communication tends to facilitate intrapersonal communication, psychotherapy may be represented as a specialized and controlled means of facilitating intrapersonal communication. Barriers between dissociated segments of the personality tend to become more permeable as one continues to talk about the issues relevant to the dissociating processes. A simplified formula may be proposed as follows: *What one can talk about indicates what he can begin to accept, and what one can accept indicates what he can begin to manage.* Thus, a step is taken in the direction of a more adequate degree of self-regulation.

FREE ASSOCIATION

The fundamental rule in traditional psychoanalytic procedures is that the analysand should say or represent verbally all that comes to mind, regardless of its nature. It can be seen immediately that, in terms of the regulatory theory of consciousness, the analyst attempts to utilize a basic device for facilitating the reintroduction of segments of experience to the conscious field. Verbalization heightens awareness, and the process of talking with reduced restriction encourages exploration and discovery of significant neglected and rejected aspects of experience. There are increased chances that a rejected aspect of experience reintroduced to consciousness will enter into mutually modifying relations with other processes currently available to the field. This can provide the discovered or uncovered aspect of experience with new context and a revised meaning. Hence, some increased degree of flexibility has been achieved. However, the invitation to a patient to use free association can be extremely threatening. It is an invitation in many instances to disturb, even to disrupt, a kind of equilibrium by which the individual has become accustomed to live. Insistence on the observance of the "fundamental rule" often places the analyst and the analysand in conflict, since the defense system of the analysand is invariably activated. Traditionally, the defense mechanisms were viewed as more or less nuisance devices that prevented the analysis from progressing smoothly toward an early successful conclusion. The approach of "Off with his defenses" often has subtle similarities to the Red Queen's command of "Off with their heads!"

No therapy is adequate that does not respect the individual's defense system as a basic aspect of his personality. Free association, be-

cause of its potencies, is both a useful and a hazardous technique. To invite the individual to free associate is to invite him to lay aside his defenses and accept back into awareness that which he has found habitually unacceptable and stressful. In terms of stress-defense dynamics, free association may be either harmful or beneficial. It is hypothesized, in terms of the present theory, that free association is therapeutically successful under the following circumstances: (*a*) if the individual's defenses operate well enough so that stressful material is fed back to the conscious field at a rate that does not exceed ego strength as here defined; (*b*) if the relationship with the therapist is sufficiently supportive so that there is some increase in ego strength; and (*c*) if the therapist's interpretations do not release anxiety that exceeds the patient's stress-tolerance threshold. Free association can be less than successful, or even harmful, if the patient's secondary defenses are breached before there has occurred sufficient development of ego strength so that secondary defenses may be less necessary to the patient. If the stress of free association in conjunction with too penetrating interpretations robs the individual too quickly of his secondary defense components, he will revert to primary defense. The ensuing results depend upon the seriousness of the problem. The patient may in some of the less serious instances simply leave therapy with less accessibility to future attempts, or, if he stays in therapy, he may show psychotic trends. In the more serious instances, a psychotic episode may occur. It is hypothesized that, were adequate records of psychotic breaks during analyses available, it would be found (*a*) that the patient initially had an inadequate secondary defense elaboration and depended chiefly on primary defenses—a dynamic arrangement which was not clearly recognized by the therapist—or (*b*) that what secondary defenses he had were interpreted away before ego strength had improved sufficiently to tolerate the added stress.

The couch and a reclining position are usually thought of as related to the technique of free association. Placing the patient on a couch where he cannot face his therapist can be damaging if the initial ego strength is not high enough to accept the stress involved in such a nonsupportive relationship. For a patient with weak ego strength, which tends to go with predispositions to utilize primary defense, to be on a couch with an aloof analyst out of sight and to be requested to free associate about his problems is to take a long step in the direction of precipitating a psychotic episode.

Free association is a powerful technique for facilitating the reacquisition of experiences, feelings, and meanings to the conscious regulatory field, but the speed of reacquisition must be compatible with the balance between stress-tolerance thresholds, or thresholds of defense reaction, and current ego strength. The process traditionally referred to in psychoanalysis as "making the unconscious conscious" needs extensive review in terms of the stress-defense dynamics which are here being outlined.

INSIGHT

From what has thus far been written, it should be clear that the concept of insight can be useful, but it must be far broader in psychotherapy than the insight of nonemotional, intellectual, problem-solving behavior. Insight in psychotherapy is based on the acceptance into consciousness of hitherto neglected or rejected aspects of experience under conditions of sufficient emotional permissiveness for these aspects, as well as other relevant aspects of experience, to interact so that new and more comprehensive meanings emerge. Insights are, therefore, limited (*a*) if the boundaries to the conscious field are relatively impermeable to thoughts, feelings, and tendencies relevant to the problem, (*b*) if ego strength is low, and (*c*) if experienced stress is currently high enough to keep the primary component of the defense system active. The sudden, dramatic insights, described in the textbooks with misleading frequency, are simply one of the rarer aspects of the broader principle of intrapersonal communication referred to earlier. To be "open-minded" (the concept of the permeable barrier to the conscious field) toward one's experiences and feelings and to what these have meant to one, is a basic objective not only in psychotherapy but in the maintenance of mental health for everyone. If this condition of the conscious regulatory field is maintained, then insights may occur fairly frequently, but few of them are likely to be dramatic with regard to one's emotional life, since his adjust-

ment problems are kept up-to-date and there are no significant areas of experience that remain rejected or alienated.

While the facilitation of insightful reactions is an important goal in psychotherapy, the more fundamental objective is to provide the individual an opportunity gradually to accept and to deal realistically with aspects of himself and his experience that heretofore he has had to neglect or reject. The reaction of insight can be thought of as a criterion that such intrapersonal communication is occurring. However, insights must be more than logical; they must also represent an acceptance of the emotional meaning of experience. The logical insight by itself may indicate that the more affective aspects of experience are still being rejected. Logical or intellectual insight tends to modify behavior chiefly under nonstress conditions. Where the problem involves threat and anxiety sufficient to activate defenses, the insight must be affectively toned and comprehensive enough to indicate a real acceptance of the basic threat issue. The terms and processes of logical problem-solving are often less than adequate for behavior under stress.

If the fundamental goal of psychotherapy is to re-establish a more adequate degree of self regulation, then the techniques used must facilitate through interpersonal communication an improved intrapersonal communication. The broader concept of insight as used here becomes one criterion for movement toward this goal.

THE UNCOVERING OR INTERPRETIVE CONTINUUM

When a therapist interprets or invites a patient to explore his experiences, the procedure is sometimes referred to as an "uncovering" technique. How a therapist's response may be scaled according to the degree of uncovering represented has been outlined in previous studies (2). The concept of scaling the depth of interpretive responses of the therapist is of great importance in reference to the stress-defense dynamics. From this a clearer rationale seems to appear for determining the degree of depth of interpretive response of the therapist appropriate at any particular time for the patient. The question may be asked whether the therapist should restrict himself to listening, to repeating significant words or phrases, to clarifying or reflecting, or whether he should include greater depth in some of his responses. If the objective of therapy is aiding the patient in achieving an improvement in self-regulatory ability, then it must be assumed that he needs to accept into the conscious field and to integrate constructively into action patterns some aspects of himself and of reality around him that heretofore he has rejected or neglected. The way in which he presents his problem at any given time reveals his capacity (ego strength) to deal with the problem. What he leaves out or attempts to avoid reveals, in part, his use of primary defense (not simply resistance). How he rationalizes, compensates, and utilizes various symptom patterns reveals not only his need for secondary defense but also his capacity to blunt the sharpness of stress factors so that he can still deal with reality to some extent (4). It often occurs in the beginning of a therapy relationship that the patient is quite restricted in his approach to the problem and in what seems to him as relevant. The therapist's technique in managing the relationship is a first step in reducing stress which often aids in some coincident increase in ego strength. Material relevant to the problem, but held at a superficial level of rejection, may still not be verbalized. Stress effects are often obvious, but, if stress-tolerance thresholds are not greatly exceeded, the patient may, even so, deal with the material integratively. Any interpretation that brings in material that the patient can neither accept nor deal with constructively, that reactivates primary defenses, or that stimulates symptom formation may be said in general to exceed the therapeutically appropriate. For the therapist to insist on an interpretation which involves the risk that the patient may feel less accepted or less understood is to retard the therapeutic processes as defined here. The therapist may be as penetratingly interpretive as the current stress-defense dynamics predispose the individual to accept and to utilize such interpretations. This contention involves the assumption that the interpretation has some degree of validity that may be sifted out by continued exploration.

A persistent hazard of the more penetrating interpretation is that it can rob the patient of the ego strength-building experience of self-discovery. If the purpose of therapy is to improve the quality and adequacy of self-

regulation, then there should be maximum opportunity for the patient to make his own discoveries and his own applications. The appropriate technique for the therapist often grows out of the question put to himself, "How can I set the situation so that the patient can recognize a feeling, relationship, meaning, etc., and retain the feeling that he discovered it for himself?" While this cannot always be done, it should occur sufficiently often, or the therapist is not providing an opportunity for the self-regulatory capacities of the individual to grow. This approach not only avoids overdependence on the part of the patient, but it also represents an area in which psychotherapy and teaching should have greater overlap.

To make a sharp distinction between supportive (often called "covering") and exploratory (or "uncovering") therapy has become relatively frequent. In terms of the stress-defense dynamics which have been discussed, this kind of dichotomy is undesirable. To talk about one's problems is to uncover threatening aspects in some degree. How much revelation is appropriate depends on the relationship between therapist and patient, the degree of ego strength available to the patient, and the balance between primary and secondary segments of the defense system. The supportive-uncovering dichotomy should be thought of as a depth continuum. Currently, the therapist must depend on his clinical judgment as to how exploratory he should be or how exploratory he should allow his patient to be or at what level on the depth continuum it is appropriate to work. Depth can be defined here in terms of the typical degrees of hesitation and resistance encountered in the social group in discussing issues of dynamic importance to them. Such a scale has yet to be devised, but it is experimentally feasible and could have a variety of uses. In any case, some degree of depth is represented by any discussion. Therefore, it is important to point out that no therapist can be supportive in any adequate or relevant manner until his patient has provided some exploration of his problems. To attempt to be only "supportive" in the traditional sense is often to leave the patient with a feeling that he is not well understood or even that he has been somewhat rejected.

ABREACTION

Abreacting a traumatic or stressful experience has been recognized as having both therapeutic and nontherapeutic effects. Grinker and Spiegel (10) have said that abreactions under pentothal are not necessarily therapeutic, but may provide a beginning of useful insight. They have also said that abreactions in a bar under the influence of alcohol are nontherapeutic. Dollard and Miller (7) propose that the difference between therapeutic and nontherapeutic abreactions may be understood in terms of the concepts of a conditioned response theory of learning, e.g., reward and punishment, reinforcement and extinction, and generalization. How would stress-defense dynamics conceptualize this issue?

Abreaction, as a reliving of a past experience, provides an opportunity, under conditions of increased permeability of conscious boundaries, for a rejected segment of experience to be reintroduced to the conscious field and to be assimilated. By assimilated is meant (*a*) that the experience has an increased degree of acceptance and (*b*) that it is provided an opportunity to enter into mutually modifying relations with other processes also available to the conscious field. That is to say, in the new context it now has changed meaning. Accepting the segment of experience which has been persistently alienated implies a reduction in defense operations at the time. This can occur in spite of the stressful nature that abreaction seems at times to have only if the patient has had an increase in ego strength. Such a result grows out of the degree of confidence he has in the therapist and the treatment procedures. It is also therapeutically possible for an abreaction to be beneficent apart from a therapist or institutions, provided the individual can accept the products and deal with them with at least some degree of reduced defensiveness. This can occur only when some current set of factors has provided an increase in ego strength as defined here.

Abreaction can be nontherapeutic or even detrimental if the individual cannot accept the revived segment of experience. Revival may be brought about by means of drugs, free association, hypnosis, or other methods. Since the stress generated will activate the defense system which has maintained the rejection or

alienation, the personality may reconstitute itself along habitual lines. This may be thought of as simply a nontherapeutic result. The experience may be detrimental if the situation in which abreaction occurs is now also rejected. The therapist, the institution, or the treatment procedure may be rejected and the patient is then less accessible than he was previously. The area of trauma has become enlarged. In general, to be therapeutically successful, abreactions should occur only when there is a good relationship of confidence and mutual acceptance between therapist and patient. How soon this occurs in therapy depends on the personalities of the people involved; it may occur in the first interview or it may never occur.

THE RELATIONSHIP ASPECT OF PSYCHOTHERAPY

The importance of establishing and managing a relationship between therapist and patient that may be utilized for therapeutic ends has long been recognized. This relationship, the subject of many treatises, is often described in language vague and mystical. It is hoped that a step in the direction of a more operational approach is accomplished by invoking stress-defense dynamics in this issue. When the therapist is accepting and relatively nonjudgmental toward his patient in managing this relationship, he is (*a*) reducing a sense of stress and (*b*) reducing to some degree the patient's need for defense. With less need for defense, there is some increase in the permeability of the barriers to the conscious field. Coincidentally, the patient perceives not only the therapist's acceptance but his imperturbability or lack of surprise at the "terrible" things the patient tells him. This apparent strength of the therapist is a support to the patient's own ego strength. Thus, on this basis also, there is some increase in the patient's ability to accept into consciousness potentially stressful material with increased chances of dealing with it constructively. Not only has a better basis for interpersonal communication been laid, but some facilitation in intrapersonal communication may occur. Whole therapies are based on providing an individual with an opportunity to talk out his problems in a nonthreatening relationship. Some therapists have been so convinced of the effectiveness of the relationship itself that they have tended to renounce analytically derived free association and interpretation. One may hypothesize in terms of stress-defense dynamics that the aspects of relationship described above are basic to any additional elaboration of technique. According to Fiedler (8), the "experts" of various schools tend to have greatest similarity at the point of managing the relationship aspect of therapy.

The facets of the relationship between therapist and patient often become both subtle and complex. In good relationship therapy, the patient utilizes the relationship for a broader and more complete expression of his own rejected feelings and tendencies. The therapist utilizes the relationship to aid the patient in becoming increasingly aware of the nature of his problems. Thus, according to the regulatory theory of consciousness, to become aware under conditions of reduced threat is to increase chances of improved self-regulation.

GROUP RELATIONSHIP FACTORS

The concept of ego strength has been defined as the ability to accept into awareness potentially stressful material and to deal with it constructively. It is hypothesized further that ego strength, as here defined, is largely a product of a sense of security and a sense of belonging in a significant way to a group of other human beings. If this part of the theoretical frame of reference is accepted, several inferences are immediately derivable. Certainly one of the first steps in individual therapy is the extension of real acceptance by the therapist to the patient. Some feeling of being accepted and at least somewhat secure in the two-group relationship is necessary on the part of the patient or therapy cannot even get started.[2] An increase in group participation on the part of the patient during therapy is both diagnostic and therapeutic. This is an indication of improvement and a utilization of relationships that increase chances for further improvement.

If ego strength is a product of a sense of security and belonging, then a look at home and community relationships—our typical cultural pattern—is not very reassuring. One needs only to think of the multitude of ways

[2] Cf. Bobbitt and Clausen (1).

by which the typical community rejects, neglects, avoids, or makes invidious distinctions among its members to realize how ego-crippling experiences can hardly be avoided. The positive side of this community picture often looks good, but most communities unwittingly set situations that result in tragic damage to some of its members. This is of extreme importance since the increase of population, together with a loss of frontiers, forces people to live in closer and closer contacts. People must learn to live together constructively or they will increasingly destroy each other.

SUMMARY

Psychotherapy is defined as a set of techniques, all of which should have as their common purpose the progressive re-establishment in the individual of a more adequate degree of self-regulation. The pathologies of self-regulation have been ordered to a set of concepts referred to as stress-defense dynamics. The essential problem in psychotherapy is the recognition and appropriate manipulation of these variables in order to effect a maximum degree of reversal of pathological processes. This systematizing approach has been applied to a selected cluster of phenomena related to psychotherapy. These are as follows: communication and psychotherapy, free association, insight, effects of "uncovering" and interpretation, abreaction, the relationship aspect of psychotherapy, and group relationship factors. It is assumed that the current style of being eclectic in psychotherapy demands that certain basic concepts be systematically identified so that the recognized common factors in psychotherapy shall have an organized meaning. The regulatory theory of consciousness and the derived concepts of stress-defense dynamics are offered as a step in this direction.

REFERENCES

1. BOBBITT, J. M., & CLAUSEN, J. A. Psychotherapy and its public health implications. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald Press, 1953. Pp. 171–201.
2. COLLIER, R. M. A scale for rating the responses of the psychotherapist. *J. consult. Psychol.*, 1953, **17**, 321–326.
3. COLLIER, R. M. Outline of a theory of consciousness as a regulatory field: Preliminary statement. *J. Psychol.*, 1955, **40**, 269–274.
4. COLLIER, R. M. Consciousness as a regulatory field: A theory of psychopathology. *Psychol. Rev.*, 1956, **63**, 360–369.
5. COLLIER, R. M. A definition of treatment. *The Amer. J. Psychother.*, 1957, **11**, 45–55.
6. DEWEY, JOHN. *Psychology*. New York: Amer. Book Co., 1891, p. 2.
7. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw Hill, 1950, p. 372.
8. ELLIS, Albert. New approaches to psychotherapy techniques. *J. clin. Psychol.*, Suppl. Monogr. 1955, **11**, 208–260.
9. FIEDLER, F. E. A comparison of therapeutic relationships in psychoanalytic, non-directive, and Adlerian therapy. *J. consult. Psychol.*, 1950, **14**, 436–445.
10. GRINKER, R. R., & SPIEGEL, J. P. *Men under stress*. New York: Blakiston, 1945, p. 392.
11. LINDNER, R. M. Who shall practice psychotherapy? In M. H. Krout (Ed.), *Psychology, psychiatry, and the public interest*. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 148–160.
12. MOWRER, O. H. Learning theory and the neurotic paradox. *Amer. J. Orthopsychiat.*, 1948, **18**, 571–610.
13. SHAKOW, D. Psychology and psychiatry: A dialogue. *Amer. J. Orthopsychiat.*, 1949, **19**, 191–208; 381–396.
14. SHAW, F. J. A stimulus-response analysis of repression and insight in psychotherapy. *Psychol. Rev.*, 1946, **53**, 36–42.
15. SHOBEN, E. J. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, **46**, 366–392.
16. SHOBEN, E. J. A learning-theory interpretation of psychotherapy. *Harvard educ. Rev.*, 1948, **18**, 129–145.
17. WOODWORTH, R. S. *Psychology*. New York: Henry Holt, 1929, p. 226.

Received February 1, 1957.

THE EFFECTS OF FEAR AROUSAL AND SUPPRESSION OF FEAR UPON SOCIAL PERCEPTION

SEYMOUR FESHBACH AND ROBERT D. SINGER
University of Pennsylvania

THE PRESENT paper is concerned with some factors influencing the presence and extent of "projection" in social perception. The term "projection," of psychoanalytic origin, has accumulated diverse connotations (4, 5) and has acquired considerable explanatory power at the cost of precise analysis of the phenomena which it presumably explains. Thus, the paranoid who is defending against persecutory agents, the miser who perceives his neighbors as equally avaricious, and the fearful child whose world is full of danger and threat are all said to be "projecting."

The present experiment is delimited to the study of one of the psychological processes underlying the various forms of projection; namely, the influence of affect upon perception. Our basic proposition is as follows: when a particular affective state is aroused, there is a tendency for the affect to become connected to contemporaneous percepts and ideation. One might describe this process as the infusion of cognition with affect, a process very similar to Heinz Werner's concept of "physiognomic" perception (6). Freud (3) has also described a process of this kind and Bellak (1), in his theoretical analysis of apperceptive distortion, refers to the role of the mood of the perceiver.

Infusion is conceived of as a basic, primitive organismic process that is especially characteristic of infants and young children. It is assumed that the infusion tendency persists to some degree in adults although greatly modified by learning. One has to learn not to project, as it were; that is, learn to discriminate between one's own feelings and those of other objects in the environment. Repression that tends to reduce the effects of higher-order learning (2) should result in greater projection or infusion. If a subject is unaware of or denies his feelings, then he is unable to employ an important cue to which he has previously learned to make discriminating responses.

These considerations lead to the following specific hypotheses:

1. Under conditions of fear arousal, a subject perceives (judges) a stimulus person as significantly more fearful and anxious than under neutral affective conditions.

2. Under conditions of fear arousal, subjects encouraged to suppress recognition of their emotional reactions perceive a stimulus person as significantly more fearful and anxious than subjects encouraged to acknowledge their emotional reactions.

METHOD

Subjects

Sixty male volunteers from introductory psychology classes at a large metropolitan university participated as subjects. The *Ss* were randomly assigned to one of the following experimental treatment groups: Control (C), Fear-Expression (FE) and Fear-Suppression (FS).

Administration of Film and Fear Stimulus

The *Ss* were seen individually. Each *S* was told that the experiment dealt with the effects of distraction upon the accuracy of one's judgment of other people. They were informed that a film of a young man's performance on a number of different tasks would be shown to them, and that after presentation of the film, they would be required to judge the personality of this man. Following these instructions, electrodes were attached to the left ankle of *Ss* in the Fear groups. Electric shocks, administered in the guise of distracting stimuli, were employed to induce fear. The strength of electric shock administered to each *S* was determined by gradually varying the intensity of the shock until *S* reported that the shock was painful. The film was then presented, and eight shocks were administered at varying intervals while the *Ss* witnessed the film. At the same time, they were required to remove a set of pins from the O'Connor Finger Dexterity Board, a task that could be readily accomplished with only occasional glances at the board. A similar procedure was used for the Control group with the important exception that shock was omitted.[1]

Manipulation of Fear Expression and Fear Suppression

The instructions intended to encourage expression of fear or suppression and denial of fear were given

[1] In order to determine whether the pin removal task had any effect upon the *Ss*' reactions to the film, half the *Ss* in the Control group were not given this task to perform. The means and variances of the subsequent judgments of the two Control subgroups were practically identical, and consequently the subgroups were combined in analyzing the results.

prior to the showing of the film. The *Ss* in the Fear-Expression (FE) group were told: "Many people are disturbed by the shock. In order to perform most efficiently, the best thing to do is to be aware of and admit your feelings. If you freely express your feelings, your judgments will be more accurate." The *Ss* in the Fear-Suppression (FS) group were given contrasting instructions as follows: "Many people are disturbed by the shock. In order to perform most efficiently, the best thing to do is keep your mind off your emotional reactions and not think about them. Try to forget about your feelings and concentrate on the task; by so doing, your judgments will be more accurate."

Measurement of Experimental Effects

Subsequent to witnessing the film, *Ss* were administered a questionnaire regarding the personality characteristics of the stimulus person.[2] Five alternative choices were offered for each item in the questionnaire. The following four indices reflecting the evaluation of different personality attributes of the stimulus person were derived from the questionnaire responses:

Indirect Fear Judgment score. This index, based on 24 items, reflects the degree to which the stimulus person was judged as fearful or anxious in situations other than that depicted in the film. Sample questions: "How much tension do you think he usually works under?"; "How often do you think his sleep is fitful and disturbed?"; "In comparison to most people, how much does physical pain upset him?"

Direct Fear Judgment score. This index, based on 4 items, indicates the degree to which the stimulus person was perceived as fearful and anxious in the actual test situation depicted in the film. In making direct fear judgments, the *S* has more stimulus information available than in making indirect fear judgments. Sample questions: "How worried do you think he was while working on the various tasks?"; "How frequently did he seem to be anxious?"

Indirect Aggression Judgment score. This index, based on 13 items, reflects the degree to which the stimulus person was perceived as a generally aggressive individual. Sample questions: "How frequently do you think he loses his temper?"; "When he gets angry, how often do you think he takes it out on somebody else?"

Negative Personality Judgment score. This index, based on 15 items, reflects the extent to which unfavorable personality characteristics other than anxiety or aggressiveness were attributed to the stimulus person. Sample questions: "How much reliance do you think one can place upon his word?"; "How tolerant is he likely to be of other people's mistakes?"

A further control was introduced during the completion of the questionnaire in that the shock apparatus was detached from half of the *Ss* in each of the Fear-Arousal groups while for the remaining half, the shock apparatus was left intact and these *Ss* were given three shocks during this period. Since this condition did not result in any significant differences in subsequent judgments, these two subgroups are combined in the presentation of the results.

Apparatus

The *S* was seated at a table situated nine feet from the screen on which the film was projected. The electric shocks were administered by means of a high impedance 60-cycle inductive generator. The film, which lasted approximately 7 minutes, depicted a young man's performance on a card-sorting task, the O'Connor Wiggly-Block Test, and the O'Connor Tweezer Dexterity Test.[3] Neither sound track nor film titles were employed. The examiner administering the tests was not shown in the film except for a brief glimpse of his hands when he changed the test materials. The stimulus person, who had had previous practice with the various tasks, was instructed to complete each task as quickly as he could.

RESULTS

The means for all groups on each of the questionnaire subscales are presented in Table 1. The results of an analysis of variance of the Indirect Fear Judgments are given in Table 2. The *F* ratio is significant at less than the .01 level, the FS group having the highest Indirect Fear Judgment scores and the Control group, the lowest. The differences between each of the Fear-Arousal groups and the Control group are significant at the .01 level. The difference between the Fear-Suppression and the Fear-Expression groups, although in the predicted direction, is not statistically significant. However, examination of the individual item scores revealed a substantial difference between the two groups in the extent to which they used the most extreme fear category (alternative 5) in judging the stimulus person. As Table 3 indicates, 60 per cent of the FS group in contrast to 25 per cent of the FE group judged the stimulus person as extremely fearful on at least one of the items. This difference is significant at the .05 level of confidence. Since the differences between the Fear groups and the Control group may be taken as measures of projection, the results, in support of the first hypothesis, indicate that *Ss* experiencing shock project

[2] A copy of this questionnaire has been deposited with the American Documentation Institute, Order Document No. 5324 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C. remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] The authors wish to express their gratitude to Dr. William Nicholson who served as the principal character in the film and to the staff of the F. W. Taylor Management Laboratory of the Wharton School, University of Pennsylvania, who gave their time and facilities freely in processing the film.

their fear onto the stimulus person. There is also some support for the second hypothesis; namely, that suppression of fear facilitates the projection process.

Included in the questionnaire were four items comprising the Direct Fear Judgment Index which required *S* to judge how worried or anxious the stimulus person appeared during the test situation depicted in the film. Analysis of these data yielded results in the same direction as the previous findings, but the differences were less striking. The data, based on only a few items, were much more variable than those for Indirect Fear Judgment and, as Table 4 indicates, the *F* ratio obtained is not significant. Comparisons between groups based on each of the four items comprising the Direct Fear Judgment Index revealed no significant differences between the FE and C groups and between the FE and FS groups. However, on one of these items, the difference between the FS and C groups is significant at the .05 level, on two of the items, at the .06 level, and, on the remaining item, the difference is not significant.

The *Ss*' perception of the aggressiveness of the stimulus person was considered in the study because of the possibility of so-called complementary projection and also because the same operations which aroused fear may also have aroused aggression. An analysis of variance of the Indirect Aggression Judgment scores is presented in Table 5. The *F* ratio is significant at the .01 level, the FS group having the highest aggression perception scores and the C group the lowest. Thus, it appears that aggression as well as fear was projected in this situation.

In view of this finding, one might conjecture that the judgments of the Fear-Arousal groups are largely a reflection of a generalized negative attitude toward the experimenter and the experimental situation. If this were the case, one would expect the shocked *Ss* to be more likely than the C group to attribute negative personality characteristics (other than aggression or fear) to the stimulus person. However, an analysis of variance of the Negative Per-

TABLE 1

COMPARISON OF MEAN SUBSCALE SCORES

Subscale	Group			*t*	*t*	*t*
	Control (C) (N = 20)	Fear Expression (FE) (N = 20)	Fear Suppression (FS) (N = 20)	FE vs. C	FS vs. C	FS vs. FE
Indirect fear (24 items)	55.9	68.9	74.2	2.7**	3.9**	1.1
Direct fear (4 items)	8.5	9.7	11.2	.9	2.1*	1.2
Aggression (13 items)	30.2	34.4	38.1	1.8	3.3**	1.6
Negative Personality[a] (15 items)	47.4	44.6	46.3	—	—	—

[a] In the case of the Negative Personality Judgment items, the lower the mean, the more unfavorable the judgment.

* Significant at the .05 level.

** Significant at the .01 level.

TABLE 2

ANALYSIS OF VARIANCE OF THE INDIRECT FEAR JUDGMENTS

Source	Sum of Squares	*df*	Mean Square	*F*
Between groups	3546.55	2	1773.3	8.04**
Within groups	12573.65	57	220.6	

** Significant at the .01 level.

TABLE 3

USE OF EXTREME FEAR CATEGORIES ON INDIRECT FEAR JUDGMENT SUBSCALE

Group	Made 1 or More Extreme Fear Judgments	Made no Extreme Fear Judgments
Fear suppression (FS)	12	8
Fear expression (FE)	5	15

Note.—$\chi^2 = 5.8$; $p < .05$.

TABLE 4

ANALYSIS OF VARIANCE OF THE DIRECT FEAR JUDGMENTS

Source	Sum of Squares	*df*	Mean Square	*F*
Between groups	70.23	2	35.1	2.09
Within groups	956.10	57	16.8	($p < .20$ > .10)

TABLE 5

ANALYSIS OF VARIANCE OF THE AGGRESSION JUDGMENTS

Source	Sum of Squares	*df*	Mean Square	*F*
Between groups	633.0	2	3.7	5.76**
Within groups	3144.0	57	55.2	

** Significant at the .01 level.

TABLE 6
ANALYSIS OF VARIANCE OF THE NEGATIVE PERSONALITY JUDGMENTS

Source	Sum of Squares	*df*	Mean Square	*F*
Between groups	44.6	2	22.3	.56
Within groups	2266.8	57	39.8	

sonality Judgment scores presented in Table 6 indicates that differences among the three groups are slight and insignificant. These data indicate then that differences in perception between the experimental groups and the control group are specific to fear and aggression and cannot be attributed to a generalized "halo" effect.

While not directly relevant to the hypotheses under study, the correlation between judgments of fearfulness (Indirect) and judgments of aggressiveness is of some interest. The Pearson *r* calculated for all groups combined is .68 which is significant at the .01 level. The individual correlations for the C, FE, and FS groups are .65, .43, and .72, respectively, and are also statistically significant. The fact that even in the control group there is a substantial relationship between the tendency to perceive an individual as anxious and the tendency to perceive him as aggressive indicates that the obtained correlations are not solely a result of the experimental operations. An interpretation of this correlation as a reflection of a favorable-unfavorable attitude toward the stimulus person is suggested by the significant negative correlation of −.51 obtained in the Control group between the Indirect Fear Judgment scores and the Negative Personality Judgment scores (the lower the score on this scale, the less favorable the judgment).

DISCUSSION

The results, when considered as a whole, lend substantial support to the hypothesis that the arousal of fear results in a tendency to perceive another person as fearful and anxious, at least within the limits of the experimental operations used to arouse fear and to measure its consequences. In evaluating the significance of these experimental results, it is important to note that although all *Ss* were oriented toward making accurate personality judgments of the stimulus figure, relatively few relevant cues were available for making such judgments. In the case where more relevant cues were present, such as in those instances where *Ss* were asked to report the actual behavior or feelings of the stimulus person in the film situation, the effects of fear arousal, and particularly suppression of fear, are less striking than where indirect anxiety and fear were being judged. It has already been suggested that this difference may be a function of the difference in number of items and, hence, in the reliability of the scales used. The available data do not permit an assessment of the relative roles of statistical reliability and the constraints imposed by the stimulus, but there is sufficient theoretical and empirical basis for expecting greater perceptual distortion, or in this context, greater infusion, with reduction in stimulus information.

Of considerable interest are the results bearing upon the hypothesis that suppression of fear facilitates the tendency to project fear onto other social objects. On the measures of both Indirect Fear Judgment and Direct Fear Judgment, the Fear-Suppression group attributes a greater degree of fear and anxiety to the stimulus person than does the Fear-Expression group. When the two groups are compared with respect to the use of extreme fear categories, the difference between them is statistically reliable. These data, as well as those for Indirect Aggression Judgment, are then consistent with the hypothesis that suppression of affect facilitates projection. However, the effects of this variable are not as clear-cut as the effects of fear arousal, per se, and the results do not unequivocally support this hypothesis.

It must also be noted that the hypothesized explanation of the effects of suppression in terms of avoidance of cues essential for discrimination between the affective experience of the perceiver and the affect of the perceived object, is only one of several that can be offered to account for the experimental findings. Thus, although both experimental groups were reassured that many subjects are disturbed and upset by the experimental conditions, the Suppression group may have felt additional anxiety because of inability to suppress their feelings and thereby conform with the instructions. Also, greater conflict in this group may have contributed to a greater level of tension. Thus, the differences in amount of

projection between the Fear-Suppression and Fear-Expression groups may have been functions of differences in the amount of fear and aggression experienced. Data pertinent to the issue of the mechanism responsible for the effects of suppression could be furnished by experimental research involving positive emotions such as sympathy and joy. However, whatever the mechanism by which suppression and other variables may influence projection, it is important that one first establish the parameters of the projection or infusion process. Thus, the present results indicate that while suppression or "repression" of affect is not essential in order for projection to take place, it enhances the degree of projection present.[4]

It is instructive to examine the differences between the results of this study and those of Murray's classic study of projection (4). Although both studies deal with the effects of fear, we obtained supplementary projection and possible complementary projection while Murray obtained complementary projection only; that is, the children of Murray's study perceived the stimuli as malicious rather than fearful. An obvious explanation of this difference lies in Murray's procedure which involved rating of the pictures only with regard to maliciousness. However, we suggest that even if Murray had obtained judgments of fearfulness, changes in the ratings of fearfulness would have been less marked than changes in the ratings of maliciousness. A critical variable in our view is the relationship between the perceiver and the stimulus person. In Murray's study, the children were asked to judge pictures of adult males. Adults similar to the latter are likely to have had threatening roles in previous situations in which children have experienced fear, and consequently their faces were perceived as malicious. In terms of our theoretical framework, we would predict that if, under similar experimental conditions, pictures of children were used as stimuli, supplementary projection would be obtained; that is, the perception of the affective state of the stimulus person would directly match the affective state of the perceiver.

These considerations serve to emphasize that the final judgment or perception of a social object is the outcome of a number of interacting variables, one of which is the affective state of the perceiver.

[4] It is possible to conceive of a circumstance in which "suppression" of affect might lead to the absence of projection. Thus, if the *S* were completely successful in restricting his responses to those which were incompatible with fear responses, he might not experience any fear at all, and consequently he would not project since the present theory supposes that the response produced affective cues must be part of the stimulus field in order for projection to occur. The concept "denial" is perhaps more descriptive than are "suppression" or "inhibition" of the situation in which a particular affective response is elicited but the *S* is motivated not to recognize it as such.

SUMMARY

The present experiment was designed to study the effects of fear arousal and suppression of fear upon social perception. It was hypothesized that individuals subjected to a fear-producing situation would tend to project their feelings upon some social objects, and further, instructions to inhibit emotional reactions would increase the amount of projection.

The *S*s were randomly assigned to one of the following experimental treatment groups: Fear Expression, Fear Suppression, and Control. Shock was administered at varying intervals to *S*s in both Fear groups while they witnessed a film of a young man's performance of a number of different tasks. Prior to presentation of the film, *S*s in the Fear Expression group were encouraged to recognize any feelings they might experience because of the shock. In contrast, *S*s in the Fear Suppression group were encouraged to suppress and inhibit their emotional reactions. The Control group also witnessed the film but did not receive any shock. Subsequent to witnessing the film, all *S*s completed a questionnaire dealing with the personality characteristics of the individual in the film.

The following results were obtained:

1. The Fear-Suppression and the Fear-Expression groups perceived the stimulus person as a significantly more fearful and also significantly more aggressive individual than did the Control group.

2. The differences between the Fear-Suppression and the Fear-Expression groups were consistently in the predicted direction. The difference between these groups in their mean fear judgment scores was not statistically significant. However, a significantly greater proportion of the Fear-Suppression group in comparison to the Fear-Expression group used

extreme fear categories in describing the stimulus person.

3. The experimental differences are more striking for judgments of the fearfulness and anxiety of the stimulus person in situations other than that immediately depicted in the film.

4. The tendency to perceive the stimulus person as a fearful and anxious individual was significantly correlated with the tendency to perceive him as aggressive.

The data indicate that the arousal of fear results in a tendency to project fear onto a stimulus object in the environment. The results are also consistent with the hypothesis that suppression facilitates the tendency to project although, in this latter instance, one is less confident in rejecting the null hypothesis. Several alternative explanations of the effects of the suppression variable were considered and the role of cognitive variables in the projection process was discussed.

REFERENCES

1. BELLAK, L. On the problems of the concept of projection. In L. E. Abt & L. Bellak (Eds.), *Projective psychology*. New York: Knopf, 1952. Pp. 7–32.
2. DOLLARD, J., & MILLER, N. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
3. FREUD, S., *Totem and taboo*. London: Routledge & Kegan Paul, 1950.
4. MURRAY, H. A. The effect of fear on the estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, **4**, 310–329.
5. SEARS, R. R. Experimental studies of projection: I. Attribution of traits. *J. soc. Psychol.*, 1936, **7**, 151–163.
6. WERNER, H., *The comparative psychology of mental development*. (Rev. ed.) Chicago: Follett, 1948.

Received December 4, 1956.

THE MASCULINITY-FEMININITY DIMENSION IN NORMAL AND PSYCHOTIC SUBJECTS[1]

MAX R. REED[2]
Washington University, St. Louis

IN THE course of relating personality organization or structure to adjustment level or prognostic outcomes, the clinician must routinely attempt to assess the influence of the masculinity-femininity dimension. Psychoanalytic theory (3, 7, 8, 20) assigns a central role to this aspect of personal organization. Masculinity-femininity is viewed as a complex affair, derived from the combined effects of a number of personality determinants. Explicit relationships are postulated between genetic occurrences, body image or body feeling, and levels of ego organization (3, pp. 35–113, 418, 419). Early, severe disruptions in the sexual identification processes are thought to bring about major disturbances in body image organization (20, p. 252). The result is conceived as an admixture of cross-sex elements, a condition discordant with the body image essential to one's normal sexual role.

Such states are accounted for in psychoanalytic theory by postulating differences in levels of ego organization at the times at which early body image organization, primary identifications, and secondary identifications occur. Body image organization and primary identifications are thought to occur at a preverbal level. Fenichel (3, p. 107) suggests that primary identifications and the deepest layer of the ego are constructed through kinesthetic, olfactory, and visual sensations. On the other hand, secondary identifications and the superego are later developments, based primarily on auditory stimuli. Thus, there may be a wide discrepancy between the way a person knows that he should feel or act sexually and the way that he does feel or is capable of acting. These disparities are believed to be important contributors to a weakening of ego organization, the essential condition for psychotic illness.

The development of tests reflecting different levels of masculinity-femininity organization has paved the way for experimental investigation in this area. One line of development has made use of drawings of one type or another which, it is believed, reflect the basic masculinity-femininity organization. Franck and Rosen (6) devised a Drawing Completion Test (DC) which successfully differentiated college men and women. They reasoned that the differences in performance on this instrument reflected basic sex differences in body structure and psychosexual organization. The Draw-A-Person Test (DAP) is another drawing task believed to reflect sex differences in body image or body feeling and sexual organization. A number of studies (1, 4, 9, 11, 14, 15, 16, 22) with the DAP have yielded contradictory findings, but the test's full potential is by no means exhausted.

A second development has been the construction of verbal inventory measures which are thought to reflect the masculinity-femininity aspect of personality organization. Prominent among these measures is the MMPI Masculinity-femininity Scale (Mf), an empirically derived inventory yielding significant differences between men and women. Shepler (21) has shown that this test measures something different from the DC. While he found that both the Mf scale and the DC test differentiated men and women, he also found a low and insignificant correlation between the two test variables.

The DC, Mf, and the DAP were used in the present study to test the following hypotheses, derived from psychoanalytic theory, concerning the nature of differences in the masculinity-femininity organization of normal and psychotic women:

1. Psychotic women obtain more masculine scores on measures of body image than do normal women. (a) With the DC as a measure of body image, psychotic women yield more masculine scores than do normal women on

[1] This study is a condensation of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Washington University, St. Louis. The writer wishes to express his appreciation to Dr. Frederick H. Kanfer for his guidance in the study and for his critical evaluation in preparation of the present manuscript.

[2] Now at The Manatee-Sarasota County Guidance Center, Sarasota, Florida.

this measure. (b) With the DAP as another such measure, the relative body proportions of normal subjects' male and female drawings approximate more closely actual anatomical measures of such proportions than do the like body proportions of psychotic subjects' figure drawings.

2. The verbal statements of normal women concerning their masculinity-femininity status correspond more closely to their respective measures of body image status than do those of psychotic women. (a) Psychotic women show larger differences between their DC and Mf scores than do normal female subjects. (b) Psychotic women obtain higher (more masculine) DC scores than MMPI Mf scores, whereas normal female subjects obtain DC scores that are roughly equal to or lower (more feminine) than their corresponding Mf scores.

METHOD

The study included an initial test of the hypotheses (designated Experiment I) and a replication study employing independently chosen samples and independent scorers (designated Experiment II). Experiment II was planned to examine any refinements of the original hypotheses which Experiment I might suggest and to cross validate the findings.

Selection of Subjects

The control *Ss* for Experiment I were 22 nurses chosen randomly from one class of nurses in training at the St. Louis State Hospital. The control *Ss* for Experiment II were 28 nurses chosen in the same manner from the immediately following class of nurses at the same institution. The mean age for these *Ss* in Experiment I was 20.63 years with a standard deviation of 2.49. For Experiment II the mean age was 20.14 with a standard deviation of 1.19. All *Ss* were unmarried.

Twenty-five female psychotic *Ss* were selected for each experiment. The majority of these *Ss* for Experiment I was obtained from the St. Louis State Hospital, and for Experiment II a majority was drawn from the Elgin State Hospital at Elgin, Illinois. The psychotics were matched with the nurses on level of education, but it was not possible to match the two groups on the age variable. The mean age of the psychotic *Ss* of Experiment I was 29.28 years with a standard deviation of 3.71. In Experiment II, the mean age for the psychotic *Ss* was 28.52 years with a standard deviation of 3.54. In Experiment I, fifteen *Ss* were married and ten single. Sixteen *Ss* were married and nine *Ss* were single in Experiment II. In Experiment I, the *Ss* had been hospitalized from a minimum of one to a maximum of nine years with the mean length of hospitalization falling at 3.88 years. In Experiment II, the length of hospitalization ranged from one to ten years with the mean at 3.96 years. The psychotic *Ss* in no case exhibited central nervous system involvement, nor had any *S* undergone shock therapy of any type within three months immediately preceding test administrations.

Tests and Scoring Procedures

All *Ss* were administered the DC, the group form of the MMPI, and the usual DAP procedure.

The MMPI Mf. The published hand key was used to score the Mf scale of the MMPI for all *Ss*.

The DC. The DC protocols were all scored according to the criteria given in Franck's mimeographed manual (5). Although Franck and Rosen's (6) published scoring reliabilities on female normative groups are high, it seemed desirable to obtain interscorer reliabilities on the different populations used in the present study. When the writer and two additional scorers with graduate training in psychology scored the 25 psychotic records used in Experiment I, the resulting reliability coefficients were .79, .82, and .66. The writer's scoring was used for all statistical evaluations in Experiment I. In Experiment II, the independent scoring of one of the graduate students who participated in the first experiment was used.

In order to investigate Hypothesis 2 (b), DC standard scores were computed and the scale inverted for purposes of comparison with the MMPI Mf. Thus, for this comparison, a low DC standard score is in the feminine direction, whereas a low DC raw score is in the masculine direction.

The DAP. The following body measures were taken from each figure drawing protocol: overall height, waist height, neck height, above waist height, and arm length. On the basis of Dickinson's (2) findings that they reveal most clearly and consistently differences between males and females in actual measurements, the single measure of overall height and the ratios of arm length/waist height, above waist/waist height, and waist height/neck height were employed to test Hypothesis 1 (b). In these measures, the predicted proportional relationships for normals are as follows: (*a*) male figure drawing taller than corresponding female, (*b*) male proportions of arm length/waist height larger than the like female figure proportion, (*c*) female proportion of above waist/waist height smaller than the respective male figure proportion, and (*d*) female proportion of waist height/neck height greater than the respective male figure proportion. An additional summary measure consisted of the number of *Ss* for whom all of the differences in body proportions between the male and female figure drawings were in the normal anatomical direction.

As a check on the reliability of these measures, interscorer correlations were obtained for the DAP protocols of the psychotic *Ss* employed in Experiment I. Three scorers with graduate training in psychology measured all of these figure drawings. Product-moment correlations were calculated between the three sets of measures obtained. The resulting 30 reliability coefficients ranged from .91 to .99. Consequently, only the *E*'s measures were used in the statistical tests of hypotheses.

TABLE 1
COMPARISON OF DC SCORES FOR NURSES AND PSYCHOTICS

Item	Experiment I		Experiment II	
	Nurses	Psychotics	Nurses	Psychotics
No. of Cases	22	25	28	25
Mean[a]	23.36	15.60	20.14	15.00
SD	3.41	2.70	3.13	2.59
Combined median	18.00		17.60	
No. above median	18	5	25	6
No. below median	7	17	3	19
Chi square	8.546		8.902	
p	$< .005$		$< .005$	

[a] Higher scores are in "feminine" direction.

RESULTS

From Table 1 it is seen that the psychotic *Ss* in both experiments obtained significantly less feminine DC scores than the normal *Ss*. The nonparametric median test (18) yielded significant chi-square values ($p < .005$) for the two comparisons. (All chi-square values obtained in this study were corrected for continuity.)

While these findings strongly support the hypotheses advanced, the possibility must be considered that factors other than the experimental variable contributed to the differences in DC scores. Since it was not possible to equate control and experimental groups for age, product-moment correlations were calculated between the DC scores and the *Ss*' ages. The values for the psychotic *Ss* were .20 and .04 for Experiments I and II respectively. For the nurses, the corresponding correlations were .42 and $-.09$. All correlations in the present study were tested for significance using the r-to-z transformation (17). None of these values are significant. Thus, there seems to be little appreciable influence of age on DC scores.

Evaluation of the relationship of marital status to DC performance was limited by the nature of the samples. In Experiment I, the 10 single psychotic *Ss* obtained a mean DC score of 13.7 with a standard deviation of 2.3, whereas the 15 married *Ss* obtained a mean of 16.8 and a standard deviation of 2.2. In Experiment II, the 9 single *Ss* obtained a mean DC score of 14.3 with a standard deviation of 3.1, and the 16 married *Ss* obtained a mean of 15.3 with a standard deviation of 2.1. For Experiment I the difference is significant ($t = 3.41$, $p < .01$), but for Experiment II the difference is nonsignificant ($t = .79$, $p > .10$). Although the absolute numerical means of the married *Ss* are more feminine than the DC means of the total psychotic groups, they are nevertheless significantly more masculine than the corresponding nurses' mean scores (Experiment I, $t = 7.13$, $p < .01$; Experiment II, $t = 6.11$, $p < .01$). Thus, in this study there is no conclusive evidence for a systematic relationship between DC scores and marital status.

Length of hospitalization similarly seemed to have no relationship to DC performance (Experiment I, $r = -.12$, $p > .05$; Experiment II, $r = .30$, $p > .05$). Thus, there is no evidence that extraneous factors account for the obtained differences in DC test performance.

Two additional factors must be considered in evaluating the DC score differences obtained by the experimental and control groups. First, the nature of the test suggests the possibility that simplified responses, a result of impoverished psychotic ideation, might be the major source of the more masculine scores obtained by the experimental groups. However, a rough classification of the drawings into simplified versus several other categories revealed that only 27 per cent of the psychotic drawings scored as masculine could be considered simplified. Since 25 per cent of the drawings scored as feminine could also be classed as simplified, there is little evidence that the more masculine scores of the experimental group resulted primarily from simplified drawings.

Second, in any group of *Ss* scoring similarly on the test, several different styles of drawings are found. This finding is in agreement with the results of Franck and Rosen (6). It is possible that such variations may be in some systematic way associated with different types of feminine organization of different adaptive values. Further work with the DC, aimed at refinements in the scoring system, should add to an understanding of the relationships of masculinity-femininity structure to levels of ego strength or adaptive capacity.

The DAP results, presented in Tables 2 and 3, give less convincing support for the first hypothesis. Only three of the ten chi-square values are significant.

TABLE 2
GROUP COMPARISONS ON DAP MEASURES

Experiment I

DAP Measure	Nurses		Psychotics		Chi Square	p
	F[a]	M[b]	F[a]	M[b]		
Height	13	9	7	16	2.669	<.20
Arm length/Waist height	11	1	10	10	4.088	<.05
Above waist/Waist height	18	4	12	12	3.811	<.06
Waist height/Neck height	21	1	15	8	4.355	<.05

Experiment II

DAP Measure	Nurses		Psychotics		Chi Square	p
	F[a]	M[b]	F[a]	M[b]		
Height	16	12	8	14	2.706	.10
Arm length/Waist height	12	3	9	4	0.427	<.70
Above waist/Waist height	20	2	12	7	3.118	<.07
Waist height/Neck height	21	1	13	6	2.814	<.08

[a] Subjects producing drawings yielding body proportions in normal anatomical relationships.

[b] Subjects producing drawings distorted by cross-sex proportions.

TABLE 3
GROUP COMPARISONS ON COMBINED DAP MEASURES

Item	Experiment I		Experiment II	
	Nurses	Psychotics	Nurses	Psychotics
Number producing 100% body porportions in normal direction	11	4	14	4
Number producing less than 100% body proportions in normal direction	11	18	2	13
Chi square	3.642		5.430	
p	<.06		<.02	

Again, there was no systematic relationship between DAP measures and age. Product-moment correlations between the two variables ranged from −.44 to .31 with none of the values reaching significance. This finding is consistent with that of Lehner and Gunderson (14) that the figure drawings of women show no effects of age differences in the range from 20 to 40 years.

Inspection of Table 4 reveals no conclusive evidence for a systematic relationship between the DAP measures and other test scores. None of the correlations between the DAP measures and Mf scores are significant. Only one DAP measure yields a significant correlation with

TABLE 4
TEST INTERCORRELATIONS

Measure	Test	Experiment I		Experiment II	
		Nurses	Psychotics	Nurses	Psychotics
Waist height	DC	.15	−.13	.00	−.56**
	Mf	.04	.17	−.10	.09
Arm length	DC	.37	−.65**	−.54	.10
	Mf	−.23	.00	−.11	−.29
Neck height	DC	.08	−.20	.12	−.50**
	Mf	.07	.21	.04	−.10
Above waist	DC	−.14	.16	.00	.56**
	Mf	−.02	−.12	.08	−.08
Height	DC	.07	−.96**	.04	−.42*
	Mf	.09	.18	.00	.32
Correlation between Mf and DC		−.21	.34	.07	.06

Note.—Only measurements of female drawings appear in table.

* p <.05.

** p <.01.

TABLE 5
COMPARISON OF EXPERIMENTAL GROUPS ON DC–Mf STANDARD SCORES

Item	Experiment I		Experiment II	
	Above Md.	Below Md.	Above Md.	Below Md.
Nurses	17	5	18	10
Psychotics	7	18	8	16
Chi square[a]	8.546		4.950	
p	<.005		<.04	

[a] Chi square based on median test.

the DC scores in both experiments, and this only with the psychotic groups. Thus, there is considerable evidence in this study that the DAP is measuring something different from both the DC and Mf.

In agreement with Hypothesis 2 (a), the psychotic *S*s of both experiments obtain larger absolute difference scores (Mf standard scores minus DC standard scores). Table 5 presents the significant chi-square values obtained when the median test was applied to these differences. However, some clarification of this finding is needed. The result is due primarily to the more feminine DC scores obtained by the normal *S*s rather than to any differences in Mf scores between the two groups. The Mf contributed little in discriminating the two groups.

A significantly larger number of psychotic than normal *S*s obtained DC scores that were more masculine than their corresponding Mf

TABLE 6
GROUP COMPARISONS ON MMPI Mf SCALE
(Mf measures in raw scores)

	Experiment I		Experiment II	
	Nurses	Psychotics	Nurses	Psychotics
N	22	25	28	25
M	36.59	35.76	36.10	35.50
SD	4.69	5.73	6.82	4.80
t	1.63		1.045	
p	> .10		> .30	

TABLE 7
DISTRIBUTION OF DC AND DIFFERENCE SCORES ABOUT OPTIMAL CUTTING POINTS, COMBINING DATA FROM EXPERIMENTS I AND II

Item	DC Scores		Difference Scores	
	Above 17.5	Below 17.5	Above +5.5	Below +5.5
Nurses	43	7	26	24
Psychotics	11	39	9	41
Chi square	39.668		12.362	
p	< .001		< .001	

scores, giving strong support for Hypothesis 2 (b). Chi-square values for differences obtained in Experiments I and II, respectively, are 14.41, $p < .001$, and 4.05, $p < .05$.

From Table 6 it can be seen that the Mf alone, in agreement with the findings of Renaud (19), is of little value in discriminating between major diagnostic groups differing greatly in sexual organization, and contributes little toward differentiating the normal and psychotic groups. The findings of Gough (10) that psychotic *Ss* obtain more variable scores on the Mf scale than do normal *Ss* are not borne out. Product-moment correlations calculated between Mf and DC test scores give further evidence, in agreement with Shepler (21), that these two instruments assess different attributes. For the normal *Ss*, the correlations were −.21 and .07 for Experiments I and II respectively. For the psychotic *Ss*, the corresponding correlations were .34 and .06. For samples of this size, these values may not be considered as different from zero.

In order to evaluate better the relative efficiency of the two types of measures (DC scores and difference scores) in differentiating the criterion groups, the data were combined from the two experiments, and optimal cutting points were determined for the distributions of DC and difference scores. Table 7 shows that the DC scores alone discriminated much more effectively. The optimal DC cutting score of 17.5 misplaced 18 *Ss*. Seven nurses (14 per cent) fell on the psychotic side of the distribution. Eleven psychotic *Ss*, 22 per cent of the total psychotic group, fell in the nurses' end of the distribution. The optimal difference score (Mf standard score minus DC standard score) of +5.5 misplaced 33 *Ss*. Twenty-four nurses, 48 per cent, fell at the psychotic end of the scale while nine, or 18 per cent, of the psychotic *Ss* were misplaced. The correct selections for both tests are significantly beyond chance expectancy ($p < .001$ in both instances).

A word of caution is added here. Although some of the differences obtained between criterion groups in the present study are highly significant, considerable overlap occurs in all cases. Even though the two experiments were conducted on independently chosen groups, the samples were small and some drop is evident in the significance levels obtained in the second experiment. More extensive and refined investigation needs to be carried out in this area before these measures can be employed diagnostically or for purposes of individual prediction.

DISCUSSION

The results of this study are interpreted as being in general agreement with hypotheses derived from psychoanalytic theory (3, 7, 8, 20) concerning the nature of differences along a masculinity-femininity dimension in normal and psychotic women.

The evidence supports the hypothesis that there exist two rather clearly defined levels of functioning within this dimension. One level seems to be associated with or bound to kinesthetic or perceptual-motor facets of expression and can be thought of as falling within the psychoanalytic id structure. The second level of organization appears to be more closely associated with verbal expression, and thus seems more closely associated with cultural expectations of the superego. For the psychotic groups in this study, disturbances at the nonverbal level of organization are crucial and are indicated by markedly masculine performances on the DC. It is suggested that such a deviant organization provides a core of unacceptable impulses and a steady source of life stress which contribute to the psychotic *Ss'*

problems of adaptation and to their eventual breakdown. Extensive research is needed to lay the groundwork for a more refined and comprehensive theory concerning this aspect of human development.

Future research must also determine whether the differences in DC scores represent differences in developments in the identification process or reflect some basic somatic differences, such as differences in hormonal function, or both. In addition, the relationship of styles of sexual adjustment to general adaptive functioning, and the relationship between masculinity-femininity organization and psychosis in male populations need to be explored.

SUMMARY

The present study was designed to test hypotheses derived from psychoanalytic theory that psychotic and normal women differ in their responses to measures of different levels of masculinity-femininity organization. The DC and the DAP were employed as nonverbal measures of body image status, and the MMPI Mf Scale was administered as a verbal measure. On the DC, the psychotic women obtain more masculine scores than do normals. However, using the DAP as a measure of body image, only two of four measures yielded significant differences, and these did not hold up on cross validation. The verbal statements of normal women concerning masculinity-femininity status correspond more closely to their respective measures of body image status than do those of psychotic women.

A word of caution was added concerning the DC. Although some of the differences obtained with the test between criterion groups were highly significant, the amount of overlap was too great to allow its use for individual prediction.

Theoretical implications of the study were discussed, and areas for future research indicated.

REFERENCES

1. BARKER, A. J., MATHIS, J. K., & POWERS, C. A. Drawing characteristics of male homosexuals. *J. clin. Psychol.*, 1953, **9**, 185–188.
2. DICKINSON, R. L. *Atlas of human sex anatomy*. Baltimore: Williams & Wilkins, 1949.
3. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: W. W. Norton, 1945.
4. FISHER, S., & FISHER, RHODA L. Style of sexual adjustment in disturbed women and its expression in figure drawings. *J. Psychol.*, 1952, **34**, 169–179.
5. FRANCK, KATE. *Mimeographed manual for the Drawing Completion Test*. 1949.
6. FRANCK, KATE, & ROSEN, E. A projective test of masculinity-femininity. *J. consult. Psychol.*, 1949, **13**, 247–256.
7. FREUD, S. The passing of the Oedipus Complex. *Collected papers*, Vol. II. London: Hogarth Press, 1942.
8. FREUD, S. *The ego and the id*. London: Hogarth Press, 1947.
9. GOLDWORTH, S. A comparative study of the drawings of a man and a woman done by normal, neurotic, schizophrenic, and brain-damaged individuals. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1950.
10. GOUGH, H. G. Diagnostic patterns on the MMPI. *J. clin. Psychol.* 1946, **2**, 23–37.
11. GRONICK, S., & SMITH, L. J. Sex sequence in the Draw-A-Person Test and its relation to the MMPI Masculinity-femininity Scale. *J. consult. Psychol.*, 1953, **17**, 71–73.
12. HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. (Rev. ed.) Minneapolis: Univer. of Minnesota Press, 1944.
13. JOURARD, S. M., & SECORD, P. F. Body cathexis and the ideal female figure. *J. abnorm. soc. psychol.*, 1955, **50**, 243–246.
14. LEHNER, G. F. J., & GUNDERSON, E. K. Height relationships on the Draw-A-Person Test. *J. Pers.*, 1953, **21**, 392–399.
15. MACHOVER, KAREN. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.
16. MAINFORD, FLORENCE F. A note on the use of figure drawings in the diagnosis of sexual inversion. *J. clin. Psychol.*, 1953, **9**, 188–189.
17. MCNEMAR, Q. *Psychological statistics*. New York: John Wiley & Sons, 1955.
18. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.
19. RENAUD, H. R. Clinical correlates of the masculinity-femininity scale of the Minnesota Multiphasic Personality Inventory. Unpublished doctoral dissertation, Univer. of California, 1950.
20. SCHILDER, P. *The image and appearance of the human body*. New York: International Universities Press, 1950.
21. SHEPLER, B. F. A comparison of masculinity-femininity measures. *J. consult. Psychol.*, 1951, **15**, 484–486.
22. SMITH, ELGIE. A study of sex differentiation in drawings and verbalizations of schizophrenics. *J. clin. Psychol.*, 1943, **9**, 183–185.

Received June 8, 1957.

GALVANIC SKIN RESPONSE AS A MEASURE OF PATIENT'S REACTION TO THERAPIST'S PERMISSIVENESS[1]

JAMES E. DITTES
Yale University

Most theories of the dynamics of psychotherapy agree on the importance of an attitude manifested by the therapist toward the patient, which is variously characterized as accepting, warm, nonthreatening, or permissive. But it is not always made clear what types of behavior define this "acceptance" or what is its function within the situation of psychotherapy.

"Acceptance" or "permissiveness" may be defined as an absence of various forms of punishing behavior which a patient is accustomed to receive from other persons. The nature of such anticipated punishments differs among individuals but frequently includes such cues as these: other persons fail to pay close attention to one's words; others seem unwilling or unable to understand one's own point of view or remarks; one has difficulty making his points understood; not only do others fail to give a fair hearing, but they also insist vigorously on their own point of view, interrupting, contradicting, disregarding and not attempting to reconcile disagreement; they merely announce their position, as though one's own view isn't even worth considering. Such types of behavior are punishing, in large part, because they communicate to a person that he is regarded with low esteem and as of negligible importance. Less subtle forms of such punishment are outright criticism, rebuke, expressions of rejection, and clear indications that one is considered inadequate in some respect and should change. Sympathy and friendliness may be expressed, but in a patently forced and unconvincing manner which says to a person that he is of little enough importance that he can be dealt with in a manipulative, contrived manner and not as an equal.

Such types of behavior are functionally equivalent whether they are experienced as threats of punishment or as painful punishment in themselves, since such experiences become either threatening or painful as they become associated with clearly painful experiences, such as extreme rejection and withdrawal of important satisfactions; and whether as punishment or as threat of punishment, such cues motivate avoidance behavior. For example, the stimulus of having one's words not attended to carefully may frequently be associated with the frustration of not receiving what has been requested or with such clearly painful experiences as being ignored or otherwise denied signs of regard or affection by the same person who has half-heartedly listened to one's words. Such a cue comes to motivate avoidance whether it is regarded as itself punishing or merely as a signal threatening punishment.

To such punishment or threats of punishment, the initial response of a person is likely to be some state of reaction which may be characterized as tension, anxiety, alertness, mobilization, a state which undoubtedly has a variety of physiological components, and which comes to be followed by, and thereby lessened by, some type of behavior which serves to escape or otherwise defend against the punishment. Within small groups, for example, it has been demonstrated that cues indicating low acceptance of a person by other group members results in increased conformity by that person to norms of the group, even in the face of evidence contradicting the norms (3). Among the types of behavior which are learned in response to such threats of socially administered punishment are probably those which are of interest to a therapist as neurotic symptoms and as defensive resistance within therapy.

In therapy, "permissiveness"—the absence of such punishments and threats in the behavior of the therapist—should serve two related functions (4). First, the absence of such threatening cues should allow the patient to speak and act without eliciting in him the stress or tension and accompanying defensive avoidance

[1] This study is part of a psychotherapy research program supported by United States Public Health Service Grant M-648 and supervised by Dr. John Dollard. The data used here had been previously obtained by Dr. Dollard and Dr. Frank Auld, Jr. Appreciation is due Dr. Auld, Dr. Dollard, and Dr. George Mahl for making suggestions on the manuscript and to Dr. Auld, Mrs. Ann Byler, and Mrs. Michele Morris for serving as judges in ratings of the verbal material.

measures which are ordinarily evoked by such cues in the behavior of others. Second, whenever either the therapist's behavior or the patient's own thoughts or other behavior actually do produce cues which have been accompanied by punishment in the past and hence serve as threats, the failure of these to be reinforced by actual punishment from the therapist should serve to extinguish the effectiveness of the cue as a threat: for example, the failure to be heard clearly will no longer elicit a state of tense readiness followed by some defensive measure; it will be found not to be a signal of punishment.

In practice, it is unlikely that any behavior of the therapist, even the most permissive, completely avoids cues which are threatening to the patient. As a protective, adaptive measure, a neurotic patient has probably learned to respond to a wide variety of cues as though they were threats and has acquired a "low threshold" of perception and response to such cues.

The resistive or defensive responses elicited by any such cues during therapy are probably difficult to detect and to measure for several reasons. There is much individual difference in the forms of such reactions. Any individual has a large repertory of such responses, and very different ones may be elicited by slightly different situations. Because many forms of resistance and defense themselves are frequently punished, particularly the most overt, the patient learns to be very subtle in the resistant response; the resistant, avoidant behavior —particularly of a well-practiced defender, as most patients are—probably deviates only very slightly, by the minimum possible, from ordinary nonresistive behavior, so that the resistance itself is most likely to escape detection and consequent punishment.

A potentially more stable indicator of a patient's perception and response to cues of threat is the physiological measure of the galvanic skin response. Probably alone among physiological responses, the GSR is activated solely by the sympathetic and not the parasympathetic system and is generally characterized as indicating a state of "mobilization" or "preparation" (1, 7). It seems plausible to suppose that the GSR reflects the general tension or "anxiety" or state of readiness which is the immediate reaction to a threatening cue.

The GSR record of a patient in therapy, then, should show a relationship with the "permissiveness" of the therapist, as defined in terms of the absence of various punishments or cues threatening punishment, and measures of resistance should show an increase with decreased permissiveness and increased GSR.

METHOD

This study is based on the records of the final 43 hours of therapy with a 36-year-old, white, male, neurotic patient, seen by an experienced therapist three or four times a week. The patient previously had 30 weekly sessions with the same therapist before therapy was recessed for the summer. The patient terminated therapy at the end of the series reported here. All sessions were electrically recorded and verbatim transcripts typed. A continuous record of the patient's palmar skin conductance was made during each session with apparatus previously described (2). In one session, the 14th of the present series, recording equipment failed, and this is excluded from the study.

Estimation of Therapist's Permissiveness

Using the typed transcripts, two judges independently evaluated the remarks of the therapist, using ratings of one to four on each of four scales intended to represent aspects of permissiveness. Hours were judged in random order. The four scales were defined as follows:

Attentiveness. Patient is able to feel self continually in the therapist's thoughts. There is evidence that the therapist is listening carefully and remembering what the patient tells him, that he is not preoccupied or distracted.

Understanding. The therapist shows that he understands the patient's thoughts and feelings as the patient means and feels them. The patient has no trouble getting his points across; the therapist is not stupid or obtuse. The therapist's interventions suggest thoughts or feelings which the patient can recognize as his own and do not seem to him farfetched.

Gentleness. Therapist fails to show any of the signs of rejection or disapproval which the patient may have come to expect from others. He does not condemn, criticize, scold, judge, belittle, or use sarcasm or irony. He does not give evidence of a feeling of the patient's inferiority by urging the patient to change some present behavior, by interrupting or trying to channel the patient's speech, or by insisting on his own interpretations either openly or subtly; he speaks tentatively and offers his notions only as a possibility for the patient to consider.

Over-all acceptance. The total impact of the therapist's activity is likely to make the patient feel that he is unconditionally accepted, respected, valued, "loved," is an important person, not under judgment or threat of potential punishment, is "safe" to venture thoughts and feelings which have been punished in his previous relations with other people.

Judgments on each scale were made for the total hour and for smaller units of the therapist's remarks within each hour. A unit was defined as all the remarks of the therapist separated on the typescript by less than a page of the patient's remarks. The therapist units, in other words, were separated from each other by a page or more of the patient's speech. Most hours had from four to six such units. Agreement between the two judges in judging the whole hours with the four scales summed yielded a correlation of .69[2], and the correlation of judgments in the smaller units within hours was .61. Four principal scores were used.

Whole hour permissiveness is the sum for the four scales of judgments of the hour as a whole, with both judges' scores summed. Scores ranged from 15 to 32 with the median and mean at 21. This score was used as the measure of permissiveness in all results reported unless some other score was specifically indicated.

Initial permissiveness was measured by averaging the ratings on all therapist units in the first five pages of typescript.

Opening friendliness. The behavior of the therapist in greeting the patient and in any interaction during application of the GSR electrodes was judged by two judges with respect to "friendliness," defined as responsiveness to incidental social overtures of the patient or as initiation of such conversation. Independent ranking of 36 hours by two judges yielded a Spearman rank-order correlation of .87. The six remaining hours were those both judges eliminated from the sorting as providing insufficient basis for judging; in most of these, only a bare greeting was recorded. The average rank of the two judges was used as the score for the hour on this measure.

Change in permissiveness. Instances of marked change in permissiveness within any hour were defined as any change of four or more points (representing an average of one or more points for each of the four scales) between two successive therapist units. Nineteen such instances occurred in the 42 hours.

[2] All correlation coefficients reported in this paper are product-moment correlations except as otherwise noted.

Scoring of GSR Record

A GSR was scored whenever the continuous record of conductance met one of the following criteria: (*a*) a rise at the rate of 4000 or more ohms per second with an amplitude of at least 6500 ohms; or (*b*) a rise of at least 10,000 ohms amplitude at a rate of 1333 or more ohms per second. Such a change was of the order of 3 per cent of the basal resistance level. These criteria had been used previously (2) and were found to be easily distinguishable with the calibrated paper used; and had been used to distinguish between clear-cut abrupt rises in conductance and much smaller fluctuations.

The GSR measure primarily used in this study was the frequency of such rises for each 50-minute therapy session. This frequency was approximately normally distributed with a mean of 45 and median at 43, ranging from 11 to 90. In a randomly selected set of half the hours, this frequency score correlated .95 with the total amplitude of GSR deflections within each hour. In the same sample of hours, the GSR frequency correlated −.42 with the amount of decrease during the hour of the basal resistance level, as calculated by dividing the amount of change in ohms by the beginning resistance in ohms.

For obtaining a GSR score within any part of an hour, the rate of GSRs per minute was used. This varied between .25 and 3.2 GSRs per minute in the hours examined. The criterion for determining whether a *change in rate* occurred was a difference of .25 or more GSRs per minute between successive portions of an hour.

To measure GSR rate following any therapist intervention, GSRs were counted from five minutes after the beginning of the therapist unit until five minutes after the beginning of the next unit, and the rate per minute computed. It was reasoned that the effect of the therapist's speech on the GSR would appear only after the patient had heard a substantial amount of the unit. The time of five minutes was arbitrarily selected before the data were examined. The unit of the therapist's speech typically covered somewhat more than this period.

To measure the GSR preceding a therapist

unit, GSRs were counted from the beginning of the preceding therapist unit to the beginning of the relevant therapist unit, and the rate per minute computed. It was assumed that this period best represents the unit of patient behavior which might be suspected of affecting the behavior of the therapist in the subsequent unit.

Measures of Patient's Behavior

Embarrassing sex statements. Two judges, working independently from the transcripts, identified any statement by the patient clearly acknowledging personal sexual desires and behavior. Reliability for these judgments by these judges has previously been shown to be represented by phi coefficients of the order of .80 (2). A statement was scored as an embarrassing sex statement when selected by both judges.

Resistance. Two judges independently rated the patient's remarks as to whether they dealt with emotionally significant material or were resistive in the form of discussing less emotionally significant material, such as mechanics of therapy, trivial daily activity without emotional involvement, or another person's problems. The purpose of this measure was to check the possibility that a friendly, accepting manner by the therapist could be related to a low GSR score because it encouraged the patient to behave more as though he were in a friendly social conversation and hence discuss material less likely to involve emotions and the GSR. Ratings from 1 to 4 were made for each third of an hour and for the whole hour. The sum of these four judgments by both judges was taken as the resistance score. This score ranged between 8 and 24 with a mean of 17 and median at 18. Agreement between two judges yielded a correlation coefficient of .40.

Interruptions in patient's speech. As another possible indication of the kind of resistance which might be encouraged by a friendly therapist, a count was made of the number of times in each hour that the patient paused in his speech for five seconds or more. It was assumed that more fluent, less interrupted speech might accompany more sociable and less emotionally significant remarks and that more frequent pauses might accompany more significant material. Although using a slightly different indication of fluency and silence, Mahl (5) has demonstrated that increased silence is associated with increased anxiety and that decreased silence is apparently an indication that other resistances than silence are operating. In the present study, the correlation of this score, number of pauses, with the judgment of resistance described above is $-.29$ ($p = .06$), i.e, higher resistance is associated with fewer pauses.

The frequency of pauses was also used in the analysis as a possible indication of depression in the patient.

RESULTS

Relation between Permissiveness and GSR

Each of the indicators of therapist's permissiveness showed a significant inverse relationship with frequency of galvanic skin response. Table 1 reports product-moment correlations between the measures.

Since it was also found that the GSR is related to the emotional significance of the patient's speech, as is reported in more detail below, this relationship partially obscures the relationship between therapist's permissiveness and GSR. If the emotional significance of the patient's remarks is made more homogeneous by eliminating the hours most extreme in resistance and those most extreme in emotional significance of content, the relationship between permissiveness and GSR becomes sharper. The last two columns of Table 1 report the correlation between permissiveness and GSR, disregarding, in one column, the hours which represent the top and bottom eighths on the distribution of resistance scores, and in the other column, the hours which represent the top and bottom eighths on the distribution of pause scores.

Changes in permissiveness from one hour to

TABLE 1

RELATIONSHIP BETWEEN PERMISSIVENESS AND GSR

Measure of Permissiveness	Product-moment Correlations*		
	All Hours	Excluding Extreme Hours on	
		Resistance	Pauses
Whole hour permissiveness	−.51	−.65	−.73
Initial permissiveness	−.41	−.44	−.49
Opening friendliness	−.52	−.58	−.59

* All relationships are significant at beyond the .01 level.

the next were accompanied by expected changes in the frequency of GSR. Fourteeen hours showed a decrease of two or more points in permissiveness from the next previous hour. Nine of these showed an increase in frequency of GSR, four a decrease, and one no change. Of 14 hours showing an increase in permissiveness of two or more points, three increased in GSR frequency and 11 decreased. This relationship yields a chi square of 7.0 ($p < .01$). This analysis disregards the three hours which, because of vacation or illness, followed the previous hour by a lapse of a week or more.

Eighteen embarrassing sex statements occurred, distributed among eight of the hours. A significant tendency was found for the ESS to be accompanied by a GSR in hours scored low on permissiveness and not to be accompanied by a GSR in more permissive hours.

Of 10 ESSs in four hours above the median on whole hour permissiveness, four are accompanied by a GSR and six are not. Of eight ESSs in hours below the median on permissiveness, seven are accompanied by GSR and one is not. This yields a chi square of 4.1 ($p < .05$). If the ESSs are sorted at the median according to the permissiveness score of the therapist unit immediately preceding the ESS, the relationship shows a chi square of 5.8 ($p < .02$). If opening friendliness is used as the measure of permissiveness, the relationship yields a chi square of 7.1 ($p < .01$).

Although there is evidence that judges agree on the implicit unitizing in scoring ESSs (2), the unit of an ESS is not explicitly defined, and there accordingly exists the possibility of bias in determining whether or not a GSR accompanies a particular ESS. Therefore, an alternative analysis was made, using as an objectively defined unit of the patient's speech all his remarks occurring between pauses of five seconds or more or between remarks by the therapist. The rate of GSR per minute during speeches in which an ESS occurred was calculated. The average rate of GSRs for all the periods of ESS speech in hours below the median in permissiveness is 1.8 and for hours above the median is .8. The difference in rates between the two sets of ESS speeches is shown by a t test to be significant beyond the .05 level.

The four subscales originally intended to define permissiveness proved not to contribute equally to the relationship with GSR. The multiple correlation between the four scales and GSR is $-.55$. The interjudge reliability for each scale, the intercorrelations, and the regression weights are shown in Table 2. The gentleness scale alone correlates as well with the GSR ($r = -.51$) as does the sum of the four scales. Apparently, understanding and over-all acceptance are not related to GSR except as they comprise the factor which may be labeled as gentleness or a nonpunishing manner, the chief component of permissiveness with which the GSR is related.

TABLE 2

RELIABILITIES AND INTERCORRELATIONS OF FOUR PERMISSIVENESS SUBSCALES WITH REGRESSION WEIGHTS FOR MULTIPLE CORRELATION WITH GSR

Subscale	Correlation Between Judges	Intercorrelations			Correlation With GSR	Regression Weight With GSR
		Und.	Gen.	Acc.		
Attentiveness	.14	.57	.35	.50	−.38	.24
Understanding	.53		.48	.57	−.32	−.03
Gentleness	.51			.85	−.51	.42
Acceptance	.50				−.48	.02

Alternative Interpretations

The general interpretation of these results is that the degree of apparent permissiveness or threat in the behavior of the therapist is responded to by the patient in a way which may be measured by the GSR. Before this interpretation is discussed in more detail, several alternative interpretations will be considered with additional results reported as they pertain to each one.

It could be contended that the GSR and the therapist's manner are both consequences of some aspect of the patient's behavior. For example, it may be suspected that on occasions when the patient is depressed, the GSR reflects this fact with a lower frequency and that the therapist also responds to this depression with a more kindly, sympathetic manner, which is judged more permissive. Or perhaps the patient comes to certain hours with a particular anxiety which results in more GSRs and which also results in a resistive kind of behavior with which the therapist attempts to deal by using more radical and direct interventions, which are judged less permissive.

There is evidence, however, that variations of permissiveness within hours are followed

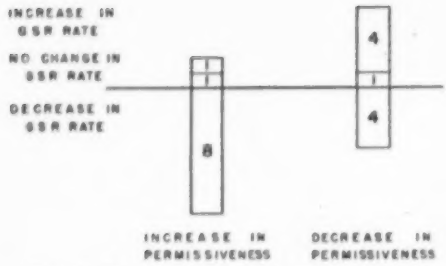

FIG. 1. EFFECT OF CHANGES IN THERAPIST'S PERMISSIVENESS ON SUBSEQUENT GSR RATE

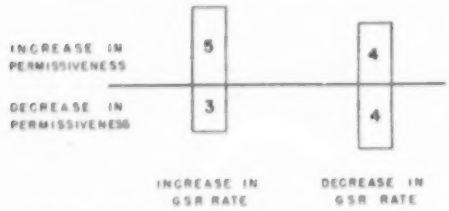

FIG. 2. EFFECT OF CHANGES IN GSR RATE ON THERAPIST'S SUBSEQUENT PERMISSIVENESS

Omitted are approximately 60 instances in which a change in the GSR rate is followed by no change in permissiveness.

but *not* preceded by corresponding inverse variations of the GSR. These data are reported in Figures 1 and 2, which record all within-hour changes of four or more points in permissiveness between two successive therapist units. The tendency of a change in permissiveness to be followed by a change in rate of GSR in the direction predicted by the interpretation held here yields a chi square of 2.7 ($p = .05$, one tail). But there is no relationship between changes in permissiveness and changes in the preceding GSR rate. These results support the interpretation that the GSR is in fact a response to the therapist's behavior.

With reference to the more specific possibilities that either depression or resistance of the patient affects both GSR and permissiveness, a possible measure of depression, frequency of pauses, though correlated with permissiveness ($r = .38$, $p < .02$) is not correlated with frequency of GSR ($r = .14$). Resistance as judged in this study is correlated inversely with permissiveness ($r = -.53$, $p < .01$) and only slightly with frequency of GSR ($r = .22$). The frequency of pauses score just reported may also be considered as an indication of resistance.

Another type of interpretation would explain the relationship of permissiveness and GSR as an artifact. It might be suspected that the therapist was judged more permissive and friendly when he talked more, thus leaving the patient less time to think and speak his own thoughts, and that listening is less likely to produce GSRs than is speaking. However, a count of the frequency of the therapist's speech (number of times therapist spoke between utterances of the patient) shows a negative correlation with permissiveness ($r = -.33$, $p < .05$) and no relation with GSR frequency ($r = .07$)

Perhaps the most likely alternative explanation is that permissiveness, as defined here, encourages certain kinds of resistance by making the patient feel that the therapy is more like a social occasion, leading him to speak of routine, less emotionally involved material; hence, fewer GSRs are to be expected. Three types of evidence are available on this point.

First, this type of resistance was judged with moderate reliability between two judges but showed little relation with GSR ($r = .22$), and furthermore yielded a negative relationship with permissiveness ($r = -.53$) contrary to that proposed by the explanation.

Second, the other measure of resistance, number of pauses by the patient, was also found related to permissiveness ($r = .38$, $p < .02$) but not to GSR ($r = .14$).

Third, variability in the patient's expression of emotionally significant material was presumably held constant in the judging of embarrassing sex statements, and as reported above, a clear relationship between permissiveness and GSR remains when such a uniform type of expression is examined. Of further suggestive importance is the fact that ESSs occur as frequently in permissive as in nonpermissive hours, contrary to what would be expected by the alternative explanation. ESSs did not, however, occur in hours of moderate permissiveness. All of the ESSs occurred in hours which rated either in the upper or lower thirds on the distribution of permissiveness.

Relation of Resistance with Permissiveness and GSR

Both measures of resistance show that in hours of moderate permissiveness (permissive-

TABLE 3
RELATION OF RESISTANCE AND GSR AT DIFFERENT LEVELS OF PERMISSIVENESS

Degree of Permissiveness	Number of Hours	Correlation of GSR with	
		Judged Resistance[a]	Frequency of Pauses
High	12	.02	.33
Moderate	15	.48	.67
Low	15	−.15	.16
All hours	42	−.22	.14

[a] Signs of the resistance judgments are reversed so that, as with the pause score, a high score represents more emotionally significant, less resistive, content.

ness scores within two points of the mean), there is a clear relationship between frequency of GSR and the extent to which the patient's speech comprises emotionally significant, nonresistive content. These correlations are reported in the second line of Table 3. Within this same group of hours, the correlation of GSR and permissiveness is slightly positive (r = .16); presumably the over-all trend of an inverse relationship between GSR and permissiveness is obscured by the strong relationship of GSR and content of patient's speech in these hours. The relation of GSR and emotionally significant content disappears when permissiveness becomes more extreme in either direction and presumably becomes the stronger determinant of the GSR as shown in the first and third lines of Table 3.

Each measure of resistance shows a significant correlation with permissiveness over all the hours. Correlation of permissiveness with judged resistance is −.53 and with pauses is .38, in each case indicating more emotionally significant, less resistive content accompanying higher permissiveness.

DISCUSSION

The behavior of a therapist who preserves, in general, a nonpunishing, warm, accepting relationship with a patient may, nevertheless, include detectable variability in this characteristic, even within a narrow range, and offers to the patient at least occasional cues which he may regard as threatening or punishing. Such cues would include any behavior which may, correctly or incorrectly, communicate to the patient that the therapist regards him as an inferior, inadequate, or otherwise unworthy and unrespected person. Apparent failure by the therapist to understand the patient's words or his interruption of the patient, or some form of scolding, or his apparent determination to force his own interpretations and points of view on the patient may be among such cues.

It appears confirmed by this study that a patient may react to such cues with a response which includes the galvanic skin response and which may represent a general tenseness and a defensive readiness to cope with the danger.

An episode occurring in one hour provides illustration and strong support for these interpretations. Hour 46, the contents of which were unknown to the investigator until near the end of the study, marks a distinctive change in the frequencies of GSR. Of the 14 preceding hours, only three have GSR frequencies under 50. Of the 28 following, only four have GSR frequencies over 50. A rank correlation (rho = .49, $p < .01$) was found between sequence and GSR frequency, but this is due almost entirely to the change at this hour. Rank correlation for hours 46 and after is −.01 and for hours before 46 is .23. (There is a slight correlation of .24 between permissiveness and sequence, but this is not significant, and there is no corresponding sharp difference at hour 46.)

About one-third of hour 46 is devoted to interpretations by the therapist concerning his apparent friendliness to the patient and to responses by the patient on this subject. In response to the patient's observation that he can't communicate as freely as he consciously wants to, the therapist remarks, "I think you have to like somebody if you want to communicate. You have to feel free. . . Who's willing to confide or tell their private thoughts to other than a friendly person?" The patient, in effect, confirms the implications of the interpretation by reporting that he has felt the therapist to be brusque and unfriendly, as evidenced by his failure to call the patient by his first name and his reluctance to engage in social banter before and after the therapy sessions. He reports his feelings during an episode in the previous day's therapy when, "I was so anxious or falling all over myself. . . trying to impress you. . . as though I were rushing in to fill a gap that seemed to be necessary to fill to maintain a relationship, for instance. . . . I say that type of thing often where I feel that I got to rush in and patch up a rupture rather than widen the rupture or act natural or expose myself."

These words of the patient seem to support the theoretical interpretation offered in this paper. In speaking of the "relationship," he presumably refers to various cues representing the therapist's regard and esteem for him, cues on which he is highly dependent. He indicates that the absence of such cues becomes a painful experience and motivates frantic efforts to defend against such a possibility. Such cues as not being called by his first name in spite of a long acquaintance and not finding a friendly response to efforts to initiate light social conversation apparently have been responded to by the patient as though they were indications of absence of esteem or some other punishing attitude, as in fact they normally would be in most social experience, and this response has included high GSR activity. Apparently, the effect of the therapist's interpretation and the patient's reply has been to let the patient realize that the therapist, unlike most other persons in his social experience, may be impersonal without being hostile or punishing. These particular cues in the therapist's behavior, though they continue to exist, lose their effectiveness as threats, and the GSR activity immediately drops and varies around a lower level.

The preliminary attempt to separate different aspects of permissiveness on separate scales results in the finding that absence of certain types of social punishment, defined here as "gentleness," proved to have the most significant relation with the GSR. In view of the unique and special emphasis sometimes given to "understanding" as the expression of a therapist's "acceptance" (6), it is of some interest to note that, as judged here and for this patient, this aspect of permissiveness is not related to the GSR independently of the scale which is most explicitly defined as the absence of punishment.

The judgment of resistance in this study was made with only moderate reliability and correlates only moderately with the objective measure of resistance, pause frequency. However, the two independent measures are highly consistent in their relationship with GSR frequency and with permissiveness and yield results that corroborate the theoretical interpretations proposed above.

If certain cues in the therapist's behavior are responded to with anxiety and fear of punishment, it would be expected that such anxiety or fear would be further increased when the patient's own behavior or speech includes characteristics that may seem more "punishable" than when his behavior or speech is more "innocent." The combination of some cues in his own behavior that have been frequently followed by punishment plus cues in the therapist's behavior that threaten punishment should evoke more anxiety (and GSR) than the latter cues alone. Personal behavior which may have been punished in some way by others would include the discussion of sexual, hostile, and other feelings toward intimately related persons—behavior, in other words, which was judged in this study as emotionally significant, less resistive.

Within the limited range of moderate permissiveness, this is precisely the relation which was found: the emotional significance of the patient's speech is significantly correlated (r = .48 or .67, depending on which measure is used) with the GSR, and is a more important determinant of the GSR than is permissiveness within these hours.

At lower levels of permissiveness, however, the findings are different. GSR is high and resistance is high (both being clearly correlated negatively with permissiveness), but the GSR and resistance are not correlated. These results suggest that anxiety (and GSR) are so uniformly high in response to the low permissiveness that they evoke generally resistive, nonintimate speech by the patient. The GSR is attributable chiefly to the low permissiveness, and not to emotionally significant content, because the latter is not present.

At high levels of permissiveness, GSR and resistance are both low, and again not significantly related with each other. Apparently, the permissiveness is great enough that the patient speaks freely of emotionally important material and does so without feeling under any threat. The cues of discussing emotionally vital matters may evoke GSR only as these cues enter into a kind of multiplicative relationship with cues of threatening behavior by the therapist. The results obtained with the ESS judgments further support these interpretations.

SUMMARY

Frequency of a patient's galvanic skin response is shown to be inversely related to the

judged permissiveness of the therapist during a series of 42 hours of psychotherapy. The GSR is interpreted as a measure of the anxiety of the patient, or his "mobilization" against any cue threatening punishment by the therapist, such as any cue which the patient has learned to perceive as evidence of low esteem. This anxiety presumably motivates varying subtle forms of resistance and defense against the feared punishment. Several alternative interpretations of the relation between permissiveness and GSR are rejected by various findings.

Two different measures of resistance are negatively correlated with the permissiveness of the therapist. However, the attempt to show that this relationship is mediated by the "anxiety" or "readiness" which the GSR measures is obscured by the finding that the GSR is related to the emotional significance of the patient's speech, as well as to the permissiveness of the therapist.

REFERENCES

1. DARROW, C. W. The galvanic skin reflex (sweating) and blood-pressure as preparatory and facilitative functions. *Psychol. Bull.*, 1936, **33**, 73–94.
2. DITTES, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, **54**, 187–191.
3. DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100–107.
4. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
5. MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1–15.
6. ROGERS, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
7. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

Received January 4, 1957.

THE INFLUENCE OF THE MOTHER'S PRESENCE ON CHILDREN'S DOLL PLAY AGGRESSION[1]

HARRY LEVIN AND VALERIE F. TURGEON
Cornell University

DURING the course of socialization, the mother teaches her child skills, beliefs, values, and brings under control the child's free expression of his impulses. Among the impulses to be controlled, aggression presents special problems. For one, the very nature of training creates frustrations that are likely to result in aggressive behavior. Such acts of aggression must in turn be controlled, a process likely to instigate more of the change-worthy behaviors, a sort of vicious circle. The socializer thereby becomes an instigator to the very action she is trying to change.

In most instances, however, mothers do succeed in bringing under control the aggressiveness they deplore. Through the punishment of aggression, the conditions that instigate aggression also instigate the anticipation of the pain that has, in the past, accompanied such actions. We expect that when the inhibitions are greater in strength than the urges to be aggressive, the child will not behave directly in the disapproved manner. The mother, then, comes to both instigate and to inhibit the child's aggressions.

Assume that the training at home is the point of origin on a continuum of stimulus generalization. Following Miller's displacement model (4), the greater the difference of a given stimulus situation from the relevant aspects of the home, the greater the probability of the overt occurrence of aggression, within the limit that aggression is still instigated. Sears and his co-workers (e.g., 2, 7, 8) have conceptualized doll play aggression in these terms. The doll play situation is assumed to be similar enough to the home to instigate aggression yet different enough to escape some of the inhibitions.

The doll play setting has certain superficial similarities to the home: a doll family, house, and furniture. The differences are also important and designed to relax inhibitions against aggression; the people are, after all, dolls, and the furniture and rooms are miniatures. Also, the experimenter is deliberately friendly and nonevaluative. The prediction has been that with inhibitions against aggression relaxed in doll play, the amount of aggression the child shows is a function of the degree of instigation, usually measured by how severely the child has been punished at home.

The present study asks how aggression in doll play is affected if the doll play setting is in some ways made more similar to the home by having the child's mother present at a doll play session. The expectation from the foregoing line of argument is that aggression should be reduced.

METHOD

Since doll play usually involves a single adult interacting with the child, a control was necessary to insure that the child's doll play behavior would be attributable to the presence of the mother and not to the circumstance that two adults were present. Consequently, the control group's doll play was observed by an adult female, previously unknown to the child. The results from this study may also be compared with the several extant sets of norms of doll play behavior in which the preschool child participated only with the adult experimenter (3, 6).

The design, then, involved two groups of children, an experimental and a control group. For each group there was a control and an experimental session. In the first session, both groups were treated alike; only the experimenter was present. In the second session, the experimental group were observed by their mothers in addition to the experimenter, and the control group by a stranger and the experimenter.

Subjects

The *Ss* were 20 children, 10 boys and 10 girls, aged three to five years. They came from upper middle-class, professional homes; 13 of the 20 were enrolled in the Cornell University Preschool. The children were first separated into groups of the same sex and age.

[1] This study was supported in part by a grant from the National Institute of Mental Health (M-901), "Pride and Shame in Children." We wish to thank Dr. Alfred L. Baldwin for his help with all aspects of the study and Drs. Urie Bronfenbrenner and John Harding for their careful readings of the manuscript.

Within each age-sex group the children were assigned randomly to the two treatments.

Procedure

The *S*s were observed in two sessions of standard, experimental doll play, each lasting 20 minutes. There was usually one day intervening between sessions, and in no case was the interval longer than three days. The doll play procedure has been described in detail elsewhere (6). A standard family of dolls was used: mother, father, boy, girl, and baby. The doll house had six uncovered rooms; the furniture was movable, but the walls were fixed. The same adult female (VT) was the experimenter in all sessions. The first session involved only the *E* and the child.

At the second session, the stranger-present (control) *S*s entered the experimental room with the *E* to find an unfamiliar female adult in the room. This person acted as an observer who did not speak either to the *E* or the *S*, but simply sat watching the session.

The mother-present *S*s entered the second session accompanied by their mothers and the *E*. The mothers had been previously instructed to watch the session but not to start interaction with the child. If the child talked to them they were to reply in a natural manner.

The sessions were observed from behind a one-way vision screen. The doll play record was unitized at the time of the observation into subject-predicate units which were recorded verbatim. "The boy runs," "the daddy spanks the girl" are examples of single units. A scoring convention somewhat different from other doll play studies was adopted in that repetitive sequences such as the "boy runs and runs and runs" was scored as one unit. From the protocols two scores were derived: (*a*) total number of units per session and (*b*) per cent aggression per session, computed as the number of aggressive units divided by the total number of units per session.

Our attempt to score the agents and objects of the aggressive acts was not profitable. Usually, the child failed to specify either the agent or the object of the act.

Aggression was defined as any thematic action that irritates, hurts, injures, punishes, frustrates, or destroys a doll or equipment. It ranged in intensity from innocuous mischief and the violation of usual household rules (the boy jumps on the bureau) to severe accidents, fights, and catastrophes (the father breaks his leg, the boy has his head sliced off, a storm breaks up the house). Verbal agression such as scolding and threats were included, as were instances in which discomfort was attributed to a doll (the boy is sick, sad, lost, etc.).

Four sessions were observed by two observers. The four protocols were combined and agreement computed according to the formula: % Agreement = number of agreements/number of agreements + number of disagreements. For total units, the agreement between observers was 90%; for aggression units, the agreement was 83%.

RESULTS

The score computed for the aggression analysis is the percentage of total units which are aggressive. This score controls for the overall level of general *verbal* thematic productivity of the child. For the various analyses, the percentage scores were normalized by use of the angular transformation, $p = \sin^2 \phi$.

To ascertain the comparability of the experimental and control groups, an analysis of variance of the aggression scores in the first session, prior to the experimental treatment, was performed. The difference in the average per cent aggression between the two treatment groups was not significant.

Since the results of the present experiment are to be compared with existing doll play norms, it is relevant to inquire whether the doll play behavior of the present *S*s in the initial session diverges significantly from that of the groups on which norms are available. From a study of 202 boys and 177 girls, Levin and Sears (3, p. 149) report that the average per cent aggression scores in the initial doll play session is 14.28 for boys, and 8.64 for girls. The mean and .05 fiducial limits for the ten boys in this study is 23.10 ± 14.14; and for the girls, 6.90 ± 4.98. While the obtained means are not as similar as might be desired, these findings are compatible with the view that doll play performance in this situation is typical of other, larger samples of children.

TABLE 1
AGGRESSION SCORES IN TWO DOLL PLAY SESSIONS
(N = 5 in each cell)

Sex	Mother Present				Stranger Present			
	Session I		Session II		Session I		Session II	
	Mean*	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Boys	16.40 ± 21.87	17.62	41.20 ± 36.48	29.39	29.80 ± 24.48	21.34	23.60 ± 20.71	16.69
Girls	6.40 ± 13.04	10.49	10.60 ± 17.49	14.08	7.40 ± 2.72	2.19	3.80 ± 7.25	5.85

* The mean is followed in each case by fiducial limits at the .05 level.

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE OF AGGRESSION SCORES IN TWO DOLL PLAY SESSIONS

Source	*df*	Mean Squares	*F*	*p*
Treatment	1	11.02	—	—
Sex	1	2666.68	9.30	<.01
Sessions	1	126.02	—	—
Treatment × Sex	1	1.77	—	—
Treatment × Sessions	1	746.50	7.66	<.02
Sex × Sessions	1	191.85	—	—
Treatment × Sex × Sessions	1	34.22	—	—
Error (b)	(16)	286.85		
Error (w)	(16)	97.41		

The questions under investigation concerning the influence of the mother's presence in the second session were approached through an analysis of variance of normalized per cent aggression scores according to the following classifications: treatment, sex of *S*, and session. Treatment, sex, and the interaction between the two involve the sums of the aggression scores for the two sessions. For these three sources of variation, the appropriate error term is the variance estimate, among *S*s, within sex and treatment (labelled "Error b" in Table 2). The differences between the scores in the two sessions enter into the main effect of "sessions" and the various interactions between sessions and the other variables. Since the aggression scores in the two sessions are correlated, the error term for testing the differences between sessions is the interaction, *S*s × sessions, within sex and treatment (labelled "Error w" in Table 2).

For the aggression scores in both sessions combined, boys are more aggressive than girls. This finding replicates a number of other studies that reported sex differences in doll play behavior (1, 3, 5, 6).

The important results for the present study are the changes from Session 1 to Session 2. As can be seen in Table 2, the session × treatment interaction is significant at about the .02 confidence level. Aggressiveness clearly increases from the first to the second session when the mother is watching the child and decreases when a stranger is present. Every one of the ten experimental children increased in aggressiveness when the mother was present. By a sign test, the probability of there being no reversals in a series of ten observations is .01. Among the children in the presence of a stranger, the aggression scores of eight decrease in the second session; two children, one boy and one girl behaved differently from the majority.

The results in the experimental and control groups may be analyzed separately. Tested against a hypothesis of no difference, the increase of aggression in the group with the mother present is statistically significant ($t = 3.34$, $p < .01$). For the control group, the decrease is not significant ($t = 1.00$, $p < .30$).

The usual occurrence in doll play with only the *E* present is that aggression increases in the second over the first session. Is the increase in the present experimental group different from what might be expected in a normal second session without mother or stranger observing? According to available norms (3, p. 149) the mean increase in per cent aggression between sessions is 7.56 for boys and 4.12 for girls. For the present *S*s, the differences between sessions and the .05 fiducial limits of these differences are 24.80 ± 30.37 and 4.20 ± 4.91 for the boys and girls in the experimental group and -6.20 ± 28.04 and -3.60 ± 5.18 for the boys and girls in the control group. The only group that differs from the scores expected on the basis of previous results is that of girls observed by a stranger, who appear to be inhibited in their aggressiveness.

DISCUSSION

The prediction derived from displacement theory that the mother's presence would result in a decrease in her child's fantasy aggression was opposite to the results actually obtained. In light of the present findings, the displacement model appears to be an oversimplification.

In the first place, the theory assumes that in the parental treatment of aggression any [illegible]tration accumulates a potential for the child to be aggressive and any punishment for aggression tends to inhibit this behavior. Instigation and inhibition with respect to aggression are viewed as nonspecific, leading under appropriate conditions to behavior that is a resultant of the instigation and inhibition. This formulation may be scrutinized both in terms of how well it fits the complexity of aggression training and as to whether doll play, as used in the present study, provides adequate conditions for manifesting the behavior.

Recent evidence having to do with young children's aggression indicates that parents treat aggression differently depending on the nature of the aggressive acts and to whom the aggressions are directed. For example, Sears, Maccoby, and Levin (9) found, in a survey of child rearing practices, that mothers tend generally to be intolerant of aggressions directed toward themselves, somewhat more tolerant of fighting and quarreling among siblings, and still more accepting of their children's fighting with playmates. These distinctions are often sharp enough for the child to discriminate among aggressions, and to learn that under some conditions aggression is not at all condoned, whereas in other situations fighting is not only permitted but even encouraged. It may be that even with the mother present, children had categorized the laboratory situation as one appropriate to express their impulses, especially since their mothers had not unqualifiedly condemned aggression. On the other hand, it is not unlikely that one of the distinctions taught children vis-à-vis aggression is that the behavior is not appropriate in front of strangers. Such training might plausibly account for the decrease in aggressiveness in the stranger-present group.

One other discrimination that many parents probably reinforced could have influenced the behavior of the *S*s in the presence of their mothers. It is quite unlikely that the children had been restrained from make-believe fighting or from having dolls and toys fight or be mischievous. Consequently, fantasy aggression may be considered by many children as approved regardless of whether or not their mothers are watching.

The above afterthoughts do not completely invalidate the displacement theory of fantasy aggression, but they argue for greater refinement than the theory now possesses. It would seem that the prediction of aggression in doll play could be improved if the fantasy situation were analyzed as to its similarity to approved or disapproved conditions of aggression. As we stated earlier, it is unfortunate that in the present study the agents and objects of aggressive acts were not specified frequently enough to warrant their separate analysis.

Two other hypotheses may be advanced to explain the fact that the presence of the mother did not restrain the children's aggressiveness. One involves the presumption that the likelihood of retaliation for aggressive actions is a powerful inhibitor of aggression. It is soon clear to the child that there is no retribution in doll play, since the mother is in an obviously immobilized position. We may speculate that had the mothers been able to discipline their children, the amount of overt aggression would have been reduced rather than increased.

A second conjecture about the increase of aggressiveness in the mother-present group arises from the nature of superego development in the child.[2] Self control is a worrisome burden for the child so that he is ready to transfer his newly acquired control to his mother when she is available. Anecdotally, we know that mothers often complain that their children are so well behaved only when they are not around. If we assume that returning impulse control to the mother is most likely for children with the strongest superegos, we may indirectly test this "transfer of superego" hypothesis with the present data.

There is some evidence that, at the preschool ages, girls have more strongly developed superegos than do boys (9),[3] so that we might expect girls to be more influenced by the mother's presence than are boys. The sex difference analysis gives little support to this hypothesis. As can be seen in Table 2, girls are no differently influenced by the experimental treatment than are the boys.

SUMMARY

Twenty children were observed in two sessions of doll play. The first session involved the child and an experimenter. During the second session, the child's mother was an audience for one group and an adult female, not previously known to the child, watched the sessions of the second group.

As is usual in doll play, boys are more aggressive than girls. Each of the ten children observed by their mothers were more aggressive in the second than in the first session.

[2] We are indebted to Dr. John W. M. Whiting for suggesting this interpretation of the findings.

[3] Questionnaires to the mothers of the children in this study asked the same questions as (9) about signs of superego development: confessions of wrongdoing, looking sheepish or guilty and admitting transgressions when asked. These data confirm the earlier results, indicating that girls show more signs of superego than do boys.

Eight of the ten children watched by a stranger decreased their aggressions in the second session. The findings are discussed in terms of modifications of the displacement theory of aggression.

REFERENCES

1. BACH, G. R. Young children's play fantasies. *Psychol. Monogr.*, 1945, **59**, No. 2.
2. HOLLENBERG, ELEANOR, & SPERRY, MARGARET. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, **1**, 32–43.
3. LEVIN, H. & SEARS, R. R. Identification with parents as a determinant of doll play aggression. *Child Develop.*, 1956, **27**, 135–153.
4. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
5. PINTLER, MARGARET H., PHILLIPS, RUTH, & SEARS, R. R. Sex differences in the projective doll play of preschool chikdren. *J. Psychol.*, 1946, **21**, 73–80.
6. SEARS, PAULINE S. Doll play agression in normal young children: influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**, No. 6.
7. SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychol.*, 1951, **6**, 476–483.
8. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.
9. SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston: Row Peterson, 1957.

Received December 26, 1956.

VERBAL CONDITIONING AS A FUNCTION OF EXPERIMENTER CHARACTERISTICS

ARNOLD BINDER, DAVID McCONNELL[1], AND NANCY A. SJOHOLM
Indiana University

A NUMBER of experiments have confirmed Greenspoon's (5, 6) findings that various stimuli can serve as reinforcements for classes of verbal responses. It has thus far been reasonably convincingly demonstrated that such categories as "plural nouns" (5, 6), "nonplural words" (5, 6), "animals" (2), "statements of attitude toward an academic issue" (8), and "first person pronouns" (4, 7, 11) form response classes, and that such experimenter responses as "Mmm-hmm" (2, 5, 6), "Good" (4, 7, 8, 11), and "expressions of agreement" (12) are effective reinforcements in leading to an increase in the frequency of subjects' responses falling in certain of these categories. It is especially significant that, with few exceptions, the subjects had been unable to verbalize the relationship existing between their own responses and the reactions of the experimenter (so-called "unconscious learning" or "learning without awareness").

Modern day interviewers of many differing schools of thought stress the importance of the interpersonal aspects of the interviewing process. According to this viewpoint, the interviewer (whether psychotherapeutic or otherwise) is an important variable to consider, since the responses of the interviewee are influenced by that particular interviewer's physical and behavioral characteristics. The changes in interviewee's expressions of opinion concomitant with alterations in the characteristics of the interviewer have been shown in the studies reported by Cantril (3) and by Robinson and Rohde (9). Sarason (10) has perhaps been most articulate in his advocacy that any complete analysis of a clinical or experimental situation involving a subject and an experimenter (or examiner or interviewer) must include consideration of the manner in which the social and personal characteristics of the latter influence the behavior of the former.

The purpose of the present experiment was to study the effect of experimenter differences upon performance in a verbal conditioning situation. The question at issue was: Would two experimenters differing markedly in physical and social characteristics produce different rates of learning in randomly selected subjects in an experiment similar to those in which minimum experimenter response has produced changes in the verbal behavior of a subject without his awareness? This experiment can, of course, only be considered preliminary since positive results would indicate the need for further research aimed at finding the specific effects of varying experimenter variables upon dependent learning variables.

The class of subject responses reinforced for this experiment was the use of "mildly hostile" words in sentences made up by the subject, and the reinforcement selected was the response "Good." The hostile word class was chosen because of the interesting incidental, interview-relevant comparison that might thereby be facilitated.

METHOD

Subjects

Two groups of *Ss* were run for the study, both selected by means of routing slips voluntarily signed in a few introductory psychology courses at Indiana University. The two groups consisted of 29 and 30 *Ss*, respectively.

Experimenters

The first group (29 *Ss*) was run by an attractive, soft-spoken, reserved, young lady who was 5′½″ in height and 90 pounds in weight. The experimenter for the second group (30 *Ss*) was very masculine, 6′5″ tall, 220 pounds in weight, and had many of the unrestrained personality characteristics which might be expected of a former marine captain. Both were students at Indiana University. Perhaps more important than their actual age difference of about 12 years was the difference in their age appearance; the young lady could have passed for a high school sophomore while the male experimenter was often mistaken for a faculty member. Henceforth, for the sake of convenience, we shall refer to the former experimenter as *FE* and the latter as *ME*.

Materials and Apparatus

Table 1 shows the two sets of words used in this experiment, one set consisting of "mildly hostile" verbs and the other of "neutral" verbs, all in the past

[1] Now at Ohio State University.

TABLE 1
VERBS PRESENTED TO SUBJECTS

Neutral	Hostile
scribbled	harmed
snored	raged
saluted	sneered
paved	squabbled
heard	vexed
lost	blamed
bounced	criticized
charted	bickered
assigned	aggravated
shrugged	punished
paraded	irritated
recited	resented
enclosed	hissed
reclined	quarreled
misplaced	annoyed
enlarged	jeered
compiled	pestered
subtracted	disliked
emerged	scolded
regulated	scowled

tense. The words were selected from larger lists generously supplied to the experimenters by Dr. Arnold Buss. Buss had six judges sort a large list of verbs into one of three categories: "neutral," "mildly hostile," and "intensely hostile." A word was accepted in a particular class only if five of the six judges agreed on its category. No "intensely hostile" words were used in the present study.

Seven sets of 20 pairs, each pair consisting of one hostile word and one neutral word, were obtained by a process of random matching. The matching of hostile words with neutral words for each set of pairs was done without replacement so that the 20 pairs in each set contained every word. Each of the pairs was typed in capital letters on a 3" x 5" white card. The pairs were placed about 1½" from the bottom of each card, both words on a single line, and roughly centered with a blank space between them. The right or left position of hostile and neutral words was randomized.

About ¾" above the paired verbs on each card, again in a single line of type, were placed the three pronouns "she," "he," and "they" in capital letters. The order of pronouns on each card was randomly determined, although the same three were used on all cards.

The presentation apparatus consisted simply of a front board (10½" in height by 18½" in width) and two side boards (10½" x 19") attached to the front board and extending back to the left and right of the experimenter. The purposes of these boards were to conceal the activities of the experimenter and to support a slotted holder for the 3" x 5" cards on the upper center portion of the front board. The apparatus was placed on a table of such height that even the shortest *S* could see the head of the shorter experimenter, *FE*.

Procedure

Before starting the trials of each new *S*, the experimenter shuffled each of the seven stacks of cards which were before him, but concealed from *S* by the boards described above. There was a sample card in the slotted holder as each *S* sat down in the experimental room, with the verbs "walked" and "ran" typed in the lower portion and the pronouns "he," "they," and "she" in the upper portion. The following instructions were then read:

> Sit in front of this stand in such a manner that you can easily read the words on the card before you. I am going to put other cards before you one by one. Each of these will be like the one before you with five words on it, three words in the upper part of the card and two in the lower part. For each card, I would like you to make up a sentence using one of the three upper words as the first word in the sentence and one of the lower two words as the second word in the sentence. You may use any one of the three upper words and either one of the two lower words in your sentence.
>
> I will place a new card before you each time you make up a sentence using the words of the preceding card. Now, it doesn't matter whether the sentence you make up is long or short, or even if it's simple or complex. It is important that you respond promptly with the first sentence that occurs to you. Let me stress that I would like you to respond with the sentence which occurs to you immediately after the exposure of the word card without further analysis or selection.

After reading the instructions and answering any questions by rereading the relevant portions of the instructions, the experimenter presented the first stack card which happened to come out on top as a result of the preliminary shuffling. In order to find operant response level, the experimenter provided no reinforcement for *S* responses to either this initial card or any of the subsequent cards in the first stack. After completing the first stack in this fashion, the experimenter presented the first card of the second stack without any delay between stacks or any indication to *S* that he had begun another stack. If the sentence made up by *S* in response to this card included the hostile verb rather than the neutral verb (regardless of the particular pronoun used), the experimenter responded, "Good." The experimenter said nothing if *S* used the neutral verb. The experimenter then proceeded through the remaining 19 cards in the second stack, reinforcing the use of hostile words with "Good" and not responding to the use of neutral words. The remaining stacks were presented in the same way without pause or change in pattern when proceeding from the last card of one stack to the first card of the next stack.

At the conclusion of the above procedure, each *S* was asked questions to determine whether or not he had insight into the relationship between his responses and the reactions of the experimenter. Great care was taken to insure that both experimenters carried out identical procedures and reinforcement patterns.

RESULTS

Each set of 20 cards was considered a block for analysis purposes, giving a total of seven blocks of trials for each *S*. The data for each *S* consisted of the number of sentences involving

TABLE 2
MEAN NUMBER OF HOSTILE WORDS USED PER BLOCK BY ALL SUBJECTS OF THE TWO EXPERIMENTS

Experimenter	No. of Subjects	Block						
		1	2	3	4	5	6	7
FE	29	8.8	9.1	10.3	10.7	11.2	11.4	11.6
ME	30	9.4	9.8	10.1	10.6	10.9	11.1	11.2

the use of hostile words given in each of the seven blocks. Table 2 contains the mean number of hostile words used in each block for the *Ss* run by *FE* and *ME*, respectively.

Before comparing the performance of the two sets of *Ss*, it was desirable to establish whether or not any learning occurred for the two groups taken separately. For this determination, one of the tests for trend recommended by Alexander (1), was used. The test used assumes a block-to-block fluctuation pattern common to all *Ss* but requires no assumption as to the equivalence of individual slopes. Table 3 summarizes the procedures applied to the data of each experimenter to test the hypothesis of no consistent trial-to-trial fluctuation or, in other words, of plateau performance throughout.

The *F*s derived from group slope are significant beyond the .001 level and indicate that the two groups tend to increase in a linear fashion. The insignificant *F*s derived from group deviations from linearity indicate that neither group shows consistent fluctuation apart from the linear trend.

But the question immediately arises: How many of the *Ss* could verbalize the relationship between their behavior and the experimenter's responses, and would there be any consistent learning trend if only those *Ss* who showed no insight were included? Actually, seven of the *Ss* in *FE*'s group and five of the *Ss* in *ME*'s group demonstrated insight by verbalizing the contingency.

Table 4 contains the results of the same trend test as described above as applied only to those *Ss* who had no insight. There were, of course, 22 such *Ss* in the group of *FE* and 25 in *ME*'s group. Again the *F* for group slope for the *Ss* of *FE* is significant beyond the .001 level. However, the corresponding *F* for the *Ss* of *ME* just misses being significant at the .05 level (an *F* of 3.92 is necessary). The *F*s derived from group deviations from linearity are once again not significant, indicating no consistent fluctuation other than linear trend. These results, together with the finding (reported below) of a highly significant over-all slope, support the hypothesis that the conditions of this experiment brought about the unverbalized learning to use hostile words rather than neutral ones in the structure of free sentences.

The primary hypothesis that the learning rates of *Ss* in verbal conditioning are a function of the personal characteristics of the experimenters, even though the participation of the experimenters is minimal and *Ss* are not aware of the experimenter response contigency, was tested by an extension of Alexander's test for trend. The extension tests the reasonableness of concluding that two groups differ in slope. Only those *Ss* who did not demonstrate insight in the questioning at the end of the experiment were used in this analysis. The results may be

TABLE 3
SUMMARY OF TREND TESTS FOR THE TWO FULL GROUPS OF SUBJECTS TAKEN SEPARATELY

Source	Experimenter *FE*				Experimenter *ME*			
	SS	*df*	MS	*F*	SS	*df*	MS	*F*
Individual deviations from estimation (error)	807	140	5.76		519	145	3.58	
Group deviations from linearity	14.0	5	2.80	.49	3.00	5	.60	.17
Between individual slopes	298	28	10.64		384	29	13.24	
Group slope	200	1	200	34.72*	82.0	1	82.0	22.90*
Between individual means	2493	28	89.04		1214	29	41.86	
Total	3812	202			2202	209		

* $p < .001$.

TABLE 4

SUMMARY OF SEPARATE TREND TESTS FOR THE TWO GROUPS OF SUBJECTS SHOWING NO INSIGHT

Source	Experimenter *FE*				Experimenter *ME*			
	SS	*df*	MS	*F*	SS	*df*	MS	*F*
Individual deviations from estimation (error)	548	105	5.22		427	120	3.56	
Group deviations from linearity	7.0	5	1.4	.27	1.0	5	.20	.06
Between individual slopes	144	21	6.86		151	24	6.29	
Group slope	113	1	113	21.65*	13.0	1	13.0	3.65
Between individual means	331	21	15.76		529	24	22.04	
Total	1143	153			1121	174		

* $p < .001$.

seen in Table 5. The *F* coming from the between group slopes is significant beyond the .01 level, in support of the hypothesis that the two groups differ in learning trend. The significant *F* derived from between group means ($p < .05$) would lead us to reject the hypothesis that the two groups do not differ in average number of hostile words. Fig. 1 shows the learning curves of these groups over the seven blocks.

The very highly significant over-all slope ($p < .001$) requires rejection of the null hypothesis that *S*s do not show unverbalized learning to use hostile words under the conditions of this experiment.

The availability of *S*s at the time of the experiment did not allow us to equate the two groups for numbers of males and females. Not only did *FE* start out with more males than *ME*, but five of the seven *S*s eliminated (for reasons of insight) from her group were females. Of the five eliminated from *ME*'s group, three were male and two were female. Thus, for the test of difference in group slope, there were six females and 16 males in the *FE*'s group, and 15 females and 10 males in *ME*'s group. In order to justify any conclusions regarding the effect of experimenter characteristics upon learning, then, we must show that the differences found between the two groups could not be accounted for purely on the basis of the differences in sex composition.

Accordingly, the entire array of *S*s without insight was divided into two groups on the basis of sex and irrespective of the particular experimenter used to get their data. Table 6 presents a summary of the Alexander test for differences in trend as applied to the comparison between female and male groups. The fact

TABLE 5

SUMMARY OF THE TEST COMPARING THE TRENDS OF THE TWO GROUPS SHOWING NO INSIGHT

Source	SS	*df*	MS	*F*
Individual deviations from estimation (error)	977	225	4.34	
Between individual slopes	293	45	6.51	
Between individual means	860	45	19.11	
Group deviations from estimation	5.0	5	1.0	
Between group slopes	30.3	1	30.3	6.98**
Between group means	19.0	1	19.0	4.38*
Over-all deviations from linearity	5.0	5	1.0	.23
Over-all slope	93.6	1	93.6	21.57***
Total	2282.9	328		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

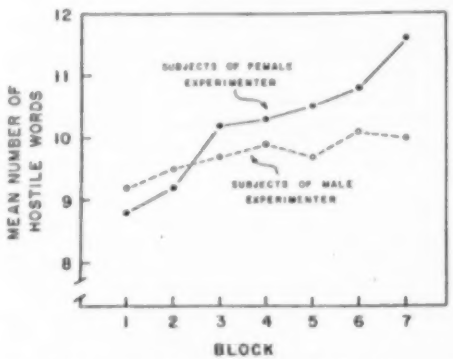

FIG. 1. LEARNING CURVES SHOWING MEAN NUMBER OF HOSTILE WORDS USED BY SUBJECTS WITHOUT INSIGHT IN THE TWO EXPERIMENTAL GROUPS

TABLE 6
SUMMARY OF THE TEST COMPARING FEMALE AND MALE SUBJECTS WITHOUT INSIGHT

Source	SS	*df*	MS	*F*
Individual deviations from estimation (error)	958	225	4.26	
Between individual slopes	319	45	7.09	
Between individual means	877	45	19.49	
Group deviations from estimation	23.0	5	4.6	
Between group slopes	3.9	1	3.9	.92
Between group means	2.0	1	2.0	.47
Over-all deviations from linearity	5.0	5	1.0	.23
Over-all slope	93.6	1	93.6	21.97*
Total	2281.5	328		

* $p < .001$.

TABLE 7
MEAN NUMBER OF HOSTILE WORDS USED PER BLOCK BY MALE AND FEMALE SUBJECTS WITHOUT INSIGHT

Sex	No. of Subjects	Block						
		1	2	3	4	5	6	7
Male	26	8.7	9.1	9.6	10.5	10.0	10.6	10.5
Female	21	9.3	9.7	10.3	9.6	10.1	10.1	11.0

that neither the difference between group slopes nor the difference between group means led to significant *F*s supports the hypothesis that males and females do not differ in learning rate or over-all level of usage of hostile words in this particular context. The mean values for the two groups may be seen in Table 7.

DISCUSSION

The principal hypothesis of this study was that learning rate in a verbal conditioning situation is a function of experimenter characteristics even though the participation of the experimenter consists only of making a single word reinforcing response and the *S* is unaware of the reinforcing contingency. This hypothesis was supported by the experimental results. However, the direction of the difference in group slope is quite interesting. As may be seen in Fig. 1, the *S*s in *FE*'s group have a lower mean number of hostile words for the first (operant) block but a higher mean number for the last block as compared with the *S*s in *ME*'s group. It surely would not have been easy to predict that the *S*s of *FE* would show a significantly greater learning rate or slope than the *S*s of *ME*. Preliminary conjecture, in fact, found it reasonable to expect that the response "Good" would have more of a reinforcing effect when uttered by the big, prestigeful, male experimenter (*ME*) than by the slightly built, soft-spoken, female experimenter (*FE*). On an a posteriori basis, however, one may speculate that *FE* provided a less threatening environment, and the *S*s consequently were less inhibited in the tendency to increase their frequency of usage of hostile words.

The finding that the experimenter response of "Good" led to a significant trend of greater usage of hostile words in sentences is of some interest in itself, as it had not been demonstrated before. Buss and Durkee[2] carried out exploratory research on the verbal conditioning of hostile words. They presented their *S*s with cards containing a "neutral," a "mildly hostile" and an "intensely hostile" verb under instructions to make up a sentence using one of the three words on each newly presented card. These investigators found that little or no conditioning was achieved when the experimenter said "Good" after correct responses and nothing after incorrect responses, although conditioning did occur when "Right" and "Wrong" were used as reinforcers. Buss and Durkee, moreover, reinforced the intensely hostile words rather than the mildly hostile words in order to maximize the chances for learning.

There were certain differences in procedure between the Buss and Durkee experiment and the present one which might account for the differing findings on the reinforcing effect of "Good." But perhaps more important than these is the fact that different experimenters were involved in the two investigations—a factor which this experiment has shown to involve significant effects.

Some of the other differences reported in the literature on verbal conditioning may be accounted for on the basis of experimenter variability. Verplanck (12), for example, reported that some of his student experimenters obtained learning effects in their *S*s while other experimenters obtained essentially plateau performance.

While it is true that the differences in learning effects found in this experiment were

[2] Buss, Arnold H., & Durkee, Ann. Personal communication. November 21, 1956.

concomitants of markedly different physical and temperamental characteristics of the experimenters, such variability is not at all uncommon in psychological experimentation. Moreover, it would not be unreasonable to expect that greater differential effects would be obtained in experiments where the extent of experimenter participation is greater than in the present experiment. Future research could profitably be directed toward determining the specific personality and/or physical characteristics of experimenters which lead to differential learning effects, the types of learning situations in which such effects occur, and the constructs from learning and personality theory which can account for the phenomena.

SUMMARY

The purpose of this investigation was to test the hypothesis that differences in the characteristics of experimenters can lead to differential learning effects when the experimental session involves verbal conditioning without awareness. Two experimenters of different sex and markedly different height, weight, age appearance, and personality ran separate groups of *S*s. The *S*s were presented with 3″ x 5″ cards which contained the pronouns "he," "she," and "they" in random order and two past tense verbs, a "neutral" one and a "mildly hostile" one. The task of each *S* was to make up a sentence starting with one of the pronouns and utilizing one of the two verbs. A total of 20 neutral and 20 hostile verbs was used, arranged in seven blocks with random pairings of neutral and hostile words within the blocks. The experimenters said "Good" whenever an *S* used the hostile verb rather than the neutral one in his freely constructed sentence.

The female experimenter's group totaled 29 *S*s, but seven of these were eliminated from the final analysis because they could verbalize the relationship between their behavior and the experimenter's response. The corresponding numbers for the male experimenter's group were 30 and five.

The data were analyzed by means of trend tests recommended by Alexander (1). The results showed that the response "Good" was reinforcing for the class of behavior consisting of the use of hostile words in sentences. In addition, it was found that the rates of learning for the *S*s of the two experimenters differed significantly, with a steeper slope for the female experimenter's group.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. BALL, R. S. Reinforcement conditioning of verbal behavior by verbal and non-verbal stimuli in a situation resembling a clinical interview. Unpublished doctoral dissertation, Indiana Univer., 1952.
3. CANTRIL, H. *Gauging public opinion*. Princeton: Princeton Univer. Press, 1944.
4. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.
5. GREENSPOON, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctoral dissertation, Indiana Univer., 1951.
6. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.
7. GROSSBERG, J. M. The effect of reinforcement schedule and response class on verbal conditioning. Unpublished doctoral dissertation, Indiana Univer., 1955.
8. HILDUM, D. C., & BROWN, R. W. Verbal reinforcement and interviewer bias. *J. abnorm. soc. Psychol.*, 1956, **53**, 108–111.
9. ROBINSON, D., & ROHDE, S. Two experiments with an anti-Semitism poll. *J. abnorm. soc. Psychol.*, 1946, **41**, 136–144.
10. SARASON, S. B. *The clinical interaction*. New York: Harper, 1954.
11. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.
12. VERPLANCK, W. S. The control of the content of conversation: reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.

Received January 18, 1957.

MEDIATED GENERALIZATION AND THE INCUBATION EFFECT AS A FUNCTION OF MANIFEST ANXIETY[1]

MARTHA T. MEDNICK[2]
Northwestern University

ONE OF the earliest and most widely cited studies attempting to verify clinically observed phenomena in the experimental laboratory was reported by Diven (6). In order to test some notions about the nature of anxiety, he conditioned a psychogalvanic response (PGR) to the word "barn" and measured autonomic reactivity to other rural words. Responsivity to the rural words was found to be greater than to a group of nonrural words, and this was considered to be evidence for the generalization of anxiety. In addition, conditioned autonomic activity tended to increase as a direct function of the length of a rest pause preceding extinction. This was called the "incubation effect." White, Cameron, and others (20, 4, 11), in their discussion of anxiety, deal with these aspects of the Diven study at length.

The Diven findings on verbal generalization may be interpreted in terms of recent attempts to explain clinical phenomena on the basis of learning theory. The study may be regarded as an instance of mediated generalization (MG), a concept discussed by Cofer and Foley (5), which may be illustrated, using Diven's verbal stimuli, as follows:

Barn + shock	elevated PGR
Barn	elevated PGR
Cow (a rural word) ... Barn	elevated PGR

"Barn," the implicit common or mediating term elicited by "cow," elicits an elevated PGR.

MG has been used in attempts to understand the role of covert symbolic processes in the determination of behavior. Dollard and Miller (7) placed considerable emphasis on this process in their discussion of personality development in terms of learning theory. These writers also used the concept of anxiety incorporated into the Hullian system by assuming that an anxiety response has drive properties and contributes to the total drive state of an organism. Some verification of this hypothesis concerning the drive properties of anxiety has been achieved through studies of the performance on learning tasks of human *S*s separated on the basis of response defined anxiety by means of extreme scores on the Taylor Manifest Anxiety Scale (MAS) (17). The results of a number of studies have supported predictions based on the assumption that MAS scores are related to drive level (18). Dollard and Miller argued further that generalization varies directly with drive and, hence, with anxiety. While stimulus generalization (generalization along a physical continuum of similarity) has been shown to vary with primary drive in animal studies (3, 12) and with anxiety (13), such a relationship has not been clearly demonstrated with MG. Lacey (9), replicating Diven's study and results, found that generalization tended to vary directly with MAS scores, but the group differences did not reach significance. The Lacey study must be interpreted with caution since the conditioning procedures utilized were somewhat irregular; the CS-UCS interval was 15 sec., filled with *S* associating aloud to the CS. This method was utilized in the interest of maintaining *S*'s unawareness of the CS-UCS contingencies.

The incubation effect described above was the other major finding reported by Diven. Bindra and Cameron (1) also studied this effect during the presentation of noxious stimulation rather than under conditioning and extinction paradigm. These investigators concluded that there appeared to be a genuine incubation effect but that little is known of its exact nature or its relation to other variables.

The present study was an attempt to reproduce Diven's and Lacey's findings with respect to the generalization and incubation of a conditioned autonomic response and to examine

[1] This paper is based upon a dissertation submitted to the Department of Psychology of Northwestern University in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to thank Prof. Janet A. Taylor, dissertation sponsor, for her advice and criticism. Mr. Bernard Bernstein gave much needed technical assistance during the pilot stages of this study. This paper was written while author was a USPHS Post-doctoral Fellow, Harvard Medical School.

[2] Now at 63 Madison Rd., Waltham, Massachusetts.

the relationship of mediated generalization to manifest anxiety. An emphasis was placed on an empirical choice of generalization stimuli. It should be noted that generalization words were chosen by Diven and Lacey on a rational basis. It was felt that additional work in this area would profit by a pre-experimental demonstration of the associational relationship of the generalization stimuli to the CS by making it possible to order the stimuli along a rough continuum and to attempt to obtain a gradient of MG.

METHOD

Subjects

Two groups of 45 *Ss* each were chosen from introductory psychology classes at Northwestern University on the basis of extreme scores on the Heineman form of the MAS (8). Forty-five individuals (15 males and 30 females) receiving scores of 64 and above on the scale were included in the high anxious group (HA), and the 45 *Ss* scoring 49 or below (35 males and 10 females) were designated as the low anxious group (LA). In addition, three LA and four HA *Ss* were tested, but because they showed no evidence of conditioning, they were excluded from further consideration. *Ss* were assigned to the various conditions by a noninvolved person so that *E* had no knowledge of MAS scores during the experiment.

Apparatus

Galvanic measures were made from a standard Stoelting polygraph. Basal readings, taken directly from a dial on the instrument, were recorded periodically. At random intervals during the presentation of the experimental stimuli, an artificial resistance of 1000 ohms was thrown into the machine. This provided a standard measuring unit for each record. The verbal stimulus materials were presented to *S* by means of a tape recorder via earphones. A loud raucous noise which served as the UCS was also presented by means of the tape recorder. A measure of the intensity of the test stimuli revealed no differences between stimuli.

Verbal Stimulus Materials

In constructing the stimulus materials, the results of a recent study by Russell and Jenkins (15) were utilized. These investigators administered the 100 words from the Kent-Rosanoff Word Association Test to 1,007 college students. A single association was elicited to each of the words and the frequency of each association determined. For the present study, the word that was chosen as the CS appeared (in varying degrees of frequency of association) as a *response* to a number of different Kent-Rosanoff stimulus words. Thus, the words "dark," "heavy," "soft," and "lamp" all had as a common response the word "light." "Light" was therefore designated as the CS in the present study and the words to which it had been associated as the generalization test words.

TABLE 1

PERCENTAGE OF INDIVIDUALS ASSOCIATING THE WORD "LIGHT" TO GENERALIZATION WORDS

Sample	*N*	Generalization Words			
		Dark	Lamp	Heavy	Soft
Minnesota	1007	83	63	58	.86
Northwestern	174	66	59	47	.1
Minnesota[a]					
HA	28	86	50	54	0
LA	28	82	61	57	0
Northwestern					
HA	25	60	60	32	4
LA	25	64	60	60	0

[a] Personal communication from Dr. Shirley Jahnson, University of Minnesota.

Table 1 presents the percentage of individuals in the Minnesota group who responded to each generalization test word with the word "light." A list of 40 of the 100 Kent-Rosanoff stimulus words containing all of the test words was also administered to a sample of Northwestern University students; the percentage of individuals responding with the word "light" to the test words in this group can also be seen in Table 1. Since it is important to show that HA and LA individuals do not differ from each other with respect to these associational frequencies, the relevant data are included in Table 1. No important differences were found between these two groups.

The stimulus materials used during the conditioning trials consisted of the CS and a list of 24 neutral words (a neutral word being one to which the CS was not given as an association by the Minnesota group). Each of the neutral words was repeated three times in differing orders, while the critical word, "light," was presented on 14 occasions. Thus, the training series consisted of 72 presentations of neutral words and 14 randomly spaced presentations of the CS. The raucous buzzer followed the CS on 9 of the 14 occasions it was presented. This partial reinforcement procedure was followed for two reasons: as an attempt to prolong extinction, and to allow the trials on which the CS was presented without the UCS to serve as a test for conditioning. The test stimulus list, used for measures of generalization and extinction, consisted of the CS, the four generalization stimuli, and three neutral stimuli. This series was repeated eight times with the order of the four generalization words and one of the neutral words ("square") being counterbalanced so as to avoid positional effects. The remaining two neutral words were included so as to increase the number of words heard by *S* between the other words. The time between the presentation of each word in both the conditioning and test series varied from 8 to 12 secs., thus allowing some time for return of *S*'s response to a basal level.

Conditioning Trials

The *Ss* in all groups received the same conditioning procedure. The experiment was conducted in a semi-soundproof room that offered relative freedom from

distraction. The earphones effectively eliminated other distracting noises. Each *S* was seated in a comfortable broad-armed easy chair with electrodes and earphones attached to him. Initial instructions, designed to acquaint *S* with the nature of the apparatus, were given by *E*. A cloth screen shielded the apparatus from *S*'s view. Additional instructions informing *S* about the stimuli to be presented were on tape and heard by *S* over the earphones. These indicated that a buzzer and words were to be presented at various intervals and that *S*'s task was merely to remain quiet and listen since his physiological reactions were being studied. No mention was made of the relationship of the buzzer to the appearance of any specific word.

Following the instructions, 15 "adaptation" words were presented to allow *S* to adjust to the apparatus before conditioning was attempted. These adaptation words were immediately followed by the conditioning series.

Test-Extinction Trials

Each of the two anxiety groups was divided at random into three subgroups for the extinction procedure. One subgroup (Condition I) underwent a regular extinction procedure immediately following the conditioning series. There was no interruption between training and extinction, no additional instructions were given, and *S*s remained attached to the apparatus. The second subgroup (Condition II) had the electrodes and the earphones removed with instructions from *E* that they would be allowed to rest; *S* remained in the chair with *E* seated out of sight behind the screen. After a lapse of 10 min., *S* was reattached to the apparatus for the extinction procedure. The *S*s were readapted to the point where deflections had mainly ceased before the tape recording was again turned on. The third subgroup of *S*s (Condition III) was allowed to leave following training and instructed to return exactly 24 hrs. later for the remainder of the experiment. Prior to his final dismissal, each *S* was asked to report whether he knew which word had been followed by the buzzer.

RESULTS

Raw ohms resistance data were obtained from the records by measuring the heights of galvanic deflections, using the 1000 ohm artificial deflection as the measuring unit. Following the method suggested by Lacey and Segal (10), ohms resistance was then transformed into conductance units, micromhos (M), by use of the formula, $M = 10^6 (1/r)$. Due to the skewed nature of the data, median measures and appropriate nonparametric statistics were used throughout the analyses. In cases where analysis of variance by ranks was used, the obtained chi-square values were converted to *F* values (19).

As noted above, there were unequal proportions of males and females in the two anxiety groups so that the anxiety and sex variables were confounded. In order to evaluate the possible effects of sex on conditioning, subgroups of 23 males and 23 females equated for MAS scores were compared as to the amount of conditioning demonstrated. The differences were not significant (chi square = 1.8). There was thus no evidence that sex was a variable influencing conditioning in this study.

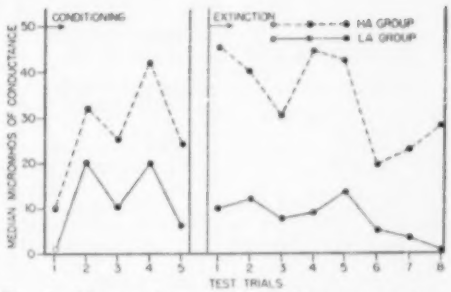

FIG. 1. MEDIAN AMPLITUDE OF CONDITIONED RESPONSES AT EACH TEST TRIAL DURING CONDITIONING AND DURING IMMEDIATE EXTINCTION (EXTINCTION DATA FOR CONDITION I ONLY)

Another preliminary analysis was made in order to see whether the three incubation groups were equally conditioned before the commencement of extinction. An analysis of variance by ranks demonstrated that these groups did not differ significantly ($F = .848$).

It should also be noted that of the 90 *S*s, 89 were able to state that "light" had been the word followed by the buzzer.

Conditioning

The responsivity to the five conditioning test trials (those occasions on which the CS was not accompanied by the buzzer) represented a measure of conditioning. As can be seen from Fig. 1, the HA group responded with greater amplitude on all test trials than did the LA group. A total conditioning score for each *S* was obtained by summing the amplitudes of all five of his conditioned responses (CRs) during training. The differences between groups in total scores, evaluated by means of chi square, were significant (chi square = 4.8, $p < .05$ for 1 *df*).

Generalization

The tests for generalization were based on the amplitude data of the first extinction series

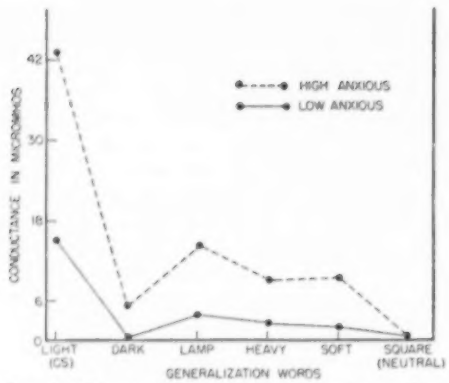

FIG. 2. MEDIAN AMPLITUDE OF RESPONSE TO EACH WORD FOR HA AND LA *Ss*

TABLE 2
CHI-SQUARE ANALYSES BETWEEN WORDS FOR ALL *Ss* COMBINED
(All at 1 *df*)

Word Pair	χ^2	p
Light-Lamp	17.28	$<.001$
Lamp-Dark	5.28	$<.05$
Lamp-Heavy	2.78	n.s.
Lamp-Soft	2.37	n.s.
Dark-Square	4.65	$<.05$
Heavy-Square	10.32	$<.01$
Soft-Square	12.00	$<.01$

because of the great increase in the number of response failures in the later series. The word "square" was used as the neutral stimulus in all comparisons since the order of presentation of this neutral word was counterbalanced. As may be seen in Fig. 2, the median amplitude for the test words tended to be higher than for the neutral word in both groups. With respect to the degree to which the test words were associated with the CS "light," the most strongly associated word "dark" elicited less responsivity than would be predicted on the basis of association alone. The remaining three stimuli roughly followed the expected pattern, with "lamp" having the greatest reactivity.

In order to test the significance of the observed differences in responsiveness between the individual words, the data for all 90 *Ss* were combined and the differences in reactivity to the test words evaluated by means of chi square. The resulting values are reported in Table 2. All the associated words elicited significantly greater reactivity than did the neutral word "square," thus demonstrating that generalization had taken place.

Turning to the relative performance of the two groups, examination of Fig. 2 reveals that the HA group showed consistently greater median amplitudes than the LA group. In order to obtain a total generalization score, the amplitudes of responses to all four test words were summed for each *S*. A chi square with respect to this overall score demonstrated the HA group to have significantly more generalization reactivity than the LA group (chi square = 9.3, $p < .05$ for 3 *df*).

Since the PGR amplitude of the HA *Ss* was significantly greater than the LA group during conditioning, this fact might be invoked to explain the obtained differences in generalization. In order to see whether these differences in generalization between the HA and LA groups could be demonstrated independently of this initial difference in conditioning level, two subgroups of 24 *Ss* each, one from the HA and one from the LA group, were matched on the basis of their total amplitude score in conditioning. A chi-square analysis between these subgroups of total generalization scores proved to be significant (chi square = 14.96, $p < .01$ for 2 *df*).

A final comparison was made among the incubation groups. These three groups, it will be recalled, all received identical training and differed only in the amount of time allowed to lapse before commencing with the regular test-extinction procedure. As previously mentioned, it was demonstrated that these groups, referred to as Condition I (no rest interval), Condition II (10-min. rest interval), and Condition III (24-hr. interval) were equally responsive at the end of training; any subsequent differences are thus attributable to the rest-interval variable. Fig. 3 summarizes the generalization data of these three conditions for the HA and LA groups combined. An analysis of variance by ranks, testing the differences between Conditions I, II, and III, was significant ($F = 15.0$, $p < .01$ for 2 and 8 *df*). Inspection of Fig. 3 shows that this significant difference may be attributed to the consistently greater amplitude of Condition III, Conditions I and II being more or less similar at each of the generalization test points. It seemed possible that the amount of incubation time might affect the HA *Ss* differently from the LA *Ss*. The same

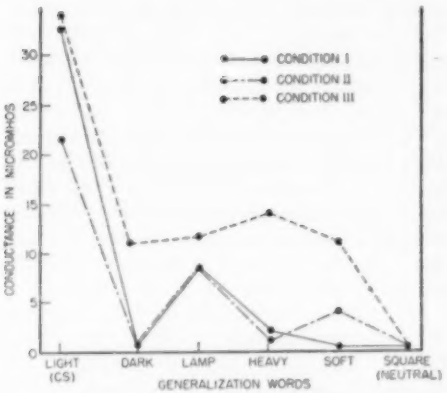

FIG. 3. MEDIAN AMPLITUDE OF RESPONSE TO EACH WORD FOR EACH CONDITION, HA AND LA *Ss* COMBINED

test between incubation conditions was therefore repeated for the LA and HA groups separately. For the HA group, the difference between Conditions I, II, and III did not reach significance ($F = .455$), but significance was reached for LA *Ss* ($F = 8.5$, $p < .05$ for 2 and 8 *df*). The significance of differences between conditions reported above could thus be traced in large part to differences in the LA group.

Extinction

The generalization data reported above represented the first of the eight series of test trials. The nature of all eight series makes it possible to follow the extinction of the CR. Only extinction of responses to the CS will be described since the test words extinguished very rapidly following the first series. Table 3 presents the median PGR amplitude at each extinction series for the CS for all groups and subgroups in the study. When the total HA and LA groups are considered, it may be seen that the HA group tended to maintain greater responsivity than the LA group until both groups finally demonstrated a median responsivity of zero. This greater responsivity was found to some degree in all three extinction conditions, but was most marked in Condition I where extinction immediately followed training. It might also be noted that contrary to the other conditions, the median PGR amplitude of the Condition I HA *Ss* did not approach zero by the end of the series, indicating that the CR did not completely extinguish.

TABLE 3
MEDIAN AMPLITUDE AT EACH EXTINCTION SERIES FOR INCUBATION AND ANXIETY SUBGROUPS

	Extinction Series							
	1	2	3	4	5	6	7	8
Cond. I								
LA	10.0	12.0	7.5	13.0	13.5	5.0	3.5	1.0
HA	45.0	40.0	30.0	34.0	42.5	19.5	23.0	28.0
Total	32.5	22.0	13.0	28.0	23.0	6.0	13.0	10.0
Cond. II								
LA	19.0	11.5	23.0	12.0	7.0	2.5	3.5	0
HA	50.0	21.0	9.5	0	9.5	0	0	0
Total	21.5	13.0	19.0	8.0	8.0	0	0	0
Cond. III								
LA	37.0	17.0	8.0	9.0	3.5	0	0	0
HA	29.0	25.5	11.5	8.0	2.5	0	0	0
Total	34.0	22.0	9.0	9.0	4.0	0	0	0
Total Group								
LA	15.0	12.5	9.0	11.0	8.0	4.5	2.0	0
HA	43.0	33.5	20.5	17.5	10.5	1.0	0	0

DISCUSSION

The results of this study confirm Diven's and Lacey's findings with respect to generalization. An inspection of the generalization data also seem to indicate that reactivity followed a rough gradient, i.e., autonomic responsivity varied directly with the degree of association of the test stimuli to the CS ("light").

Turning to the conditioning data, some comment may be made regarding the conditioning curves themselves. Although it is apparent that conditioning took place, the curves (Fig. 1) are irregular, responsivity being less at the end than in the middle of training. Such generally declining response curves have also been reported by Bitterman and Holtzman (2) and may be at least partly a function of adaptation. In the present study, an explanation in terms of the training procedure also suggested itself. It will be recalled that the CS was presented alone without the buzzer on irregularly spaced trials during training. Thus, conditioning test Trial 1 followed one UCS; two presentations of the UCS intervened between Trial 1 and Trial 2, one UCS between Trials 2 and 3, three between Trials 3 and 4, and one between Trials 4 and 5. Trials 2 and 4, consequently, had two and three UCS presentations immediately preceding them, whereas Trials 3 and 5 had only one. Inspection of Fig. 1 reveals that the irregularity is contributed to by a decline in responsivity on Trials 3 and 5. It is possible that the decline

of responsivity on these trials was at least partly related to their having fewer intervening presentations of the UCS; that is, presentation of the CS alone might have acted as an extinction trial, the effects of extinction still being noticeable when only one presentation of the UCS intervened between two test trials.

With respect to the anxiety variable, previous research using the MAS has shown that HA *Ss* exhibit both a higher level of performance in conditioning and extinction and higher gradients of SG (2, 18, 13). To the extent that SG and MG are affected by the same variables, HA *Ss* were expected to show more MG than LA *Ss*. The expected conditioning and extinction differences were found, and the expectation with respect to MG was also confirmed. It is particularly of interest, in view of the groups' differences in conditioning, that generalization differences were sustained between HA and LA subgroups equated for conditioning performance level. This supports the Dollard and Miller treatment of MG and SG as subsidiaries of the class of phenomena called generalization.

The extinction data of the group as a whole are somewhat at variance with those of other investigators (2) who reported consistently greater PGR responsivity during extinction for HA *Ss* than LA *Ss*. In the present case, this difference held up only through the fifth extinction series, after which it diminished sharply. However, it is important to note that Group I was the only group receiving immediate extinction and hence is the only one directly comparable to those of the earlier studies. An examination of Fig. 1 reveals that under Condition I, HA *Ss* showed greater responsivity throughout and also more resistance to extinction than the LA *Ss*. This result, then, supports the earlier finding. The data also suggest that an unsuspected effect of the rest interval was to speed extinction for the HA group and thus diminish the differences between the two anxiety conditions.

The incubation of anxiety has been accepted as a fairly meaningful and reliable clinical phenomenon. If an individual learns an anxiety reaction with respect to a particular stimulus situation and after a long interval returns, as it were, to the scene of the provocation, the anxiety is thought of as becoming more and more enhanced with the passage of time. As indicated, there is some experimental evidence to support the assertion of such an effect. However, the findings of this study are not completely in line with this contention. In keeping with others' results, a 24-hr. interval between training and extinction operated to increase galvanic responsivity. This effect was short-lived, however, lasting only for the first extinction series. The 10-min. interval did not result in any PGR increase. This suggests only that recovery from adaptation may be involved and that a 10-min. interval is an insufficient amount of time for the recovery to show its effect.

MG has been demonstrated to be useful in the experimental study of learning and thinking (16). The main finding of this study, the demonstration of a gradient of MG and its relationship to manifest anxiety, indicates that this may also be a useful concept in the study of pathological disturbances of thinking and learning.

SUMMARY

The present study was designed to investigate the phenomena of mediated generalization and the incubation effect. An investigation was also made of the relationship of mediated generalization to level of manifest anxiety as measured by the Taylor Manifest Anxiety Scale (MAS).

Two groups of 45 *Ss* each were chosen on the basis of extreme scores on the MAS. A conditioned PGR was first established by repeated pairings of a critical word (CS) with a raucous buzzer. Tests of mediated generalization were made following training, using as generalization stimuli, words to which the CS was associated as determined by free associations of a standardization group. In order to investigate the incubation effect the *Ss* were divided following training into three subgroups differing in the length of the rest interval between conclusion of training and the institution of extinction trials (immediate extinction, 10-min. interval, and 24-hr. interval).

The results indicated heightened responsivity towards words associated with the CS, thus indicating that mediated generalization had taken place. It was also found that extreme MAS scores were directly related to the magnitude of mediated generalization responsivity. In agreement with reports of other investigators, level of manifest anxiety was found to be directly related to conditioning

responsivity as well as to resistance to extinction.

Increased PGR reactivity was found following the 24-hr. pre-extinction interval (i.e., an incubation effect), but the influence of this effect was short-lived being confined to the early part of extinction. The 10-min. delay prior to extinction did not result in galvanic activity which was significantly greater than that found with the immediate extinction condition.

REFERENCES

1. BINDRA, D., & CAMERON, L. Changes in experimentally produced anxiety with the passage of time: incubation effect. *J. exp. Psychol.*, 1953, **45**, 197–203.
2. BITTERMAN, M. E., & HOLTZMAN, W. N. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952 **47**, 615–623.
3. BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.
4. CAMERON, N., & MAGARET, ANN. *Behavior pathology*. Cambridge, Mass.: Riverside Press, 1951.
5. COFER, N., & FOLEY, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
6. DIVEN, K. E. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.
7. DOLLARD, J., & MILLER, W. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
8. HEINEMAN, C. E. A forced choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447–454.
9. LACEY, J. I., SMITH, R. L., & GREEN, B. A. Use of conditioned autonomic responses in the study of anxiety. *Psychosom. Med.*, 1955, **17**, 208–217.
10. LACEY, O. L., & SIEGAL, P. S. An analysis of the unit of measurement of the galvanic skin response. *J. exp. Psychol.*, 1949, **39**, 122–127.
11. MURRAY, H. A. *Explorations in personality*. New York: Oxford, 1938.
12. ROSENBAUM, G. Temporal gradients of response strength with two levels of motivation. *J. exp. Psychol.*, 1951, **41**, 261–267.
13. ROSENBAUM, G. Stimulus generalization as a function of clinical and experimentally induced anxiety. Unpublished doctoral dissertation. State Univer. of Iowa, 1950.
14. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally produced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
15. RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word and Association Test. The role of language in behavior. *ONR Tech. Rep.*, No. 11, 1954.
16. RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired associate learning. The role of language in behavior. *ONR Tech. Rep.*, No. 10, 1954.
17. TAYLOR, JANET A. A personality test of manifest anxiety. *J. abnorm. soc. Psychol.* 1953, **48**, 285–290.
18. TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.* 1956, **53**, 303–320.
19. WALKER, H. M., & LEV, J. *Statistical inference*. New York: Henry Holt, 1953.
20. WHITE, R. W. *The abnormal personality*. New York: Ronald Press, 1948.

Received January 30, 1957.

A SCALE FOR MEASURING ATTITUDES OF SOCIAL RESPONSIBILITY IN CHILDREN[1]

DALE B. HARRIS

Institute of Child Welfare, University of Minnesota

EARLIER publications (3, 4) have described an attempt to evaluate an attitude of social responsibility in children. One of the scales used, adapted from the Gough responsibility scale (2), and identified as the "How I Feel About Citizenship" scale has shown virtually no increment with increasing age from 9 to 18 years. This fact, rather unique in the child psychology literature, together with the fact that a number of the items had no theoretically demonstrable relationship to the definition of social responsibility adopted for this study, suggested that the scale be revised and revalidated. For the new instrument, we retained items from the old scale and devised additional items that could be related to our theory of responsibility as a composite of attitude elements reflecting behavior classifiable as reliable, accountable, loyal, or doing an effective job. A scale of 89 items was finally prepared, each of which could be justified as representing one or more of the above mentioned attitudes. Each item was answered as "agree" or "disagree."

DEVELOPMENT OF THE SCALE

To provide a criterion, a pupil nomination form of a modified Guess Who type was administered to children in 13 seventh-and eighth-grade classes. The form asked each pupil to nominate the boys and girls who were "best fits" in his class for each of the four aspects of responsibility mentioned above.

The following untitled four descriptions, designed to illustrate these four features of responsibility, were submitted to the children:

1. This is a person upon whom you can always count; you can depend on him. When he promises that he will do something, you can always count on his doing it. You can count on his word and trust him.

2. This person is a square shooter. He doesn't try to take advantage of or cheat others. When he has done something wrong, he will own up to it rather than try to blame some one else.

3. This person thinks for the good of others rather than always for his own gain. He is loyal to the group.

4. This person is one who gets things done. On a class project, on a committee or on a work job he gets right to work and can be counted upon to do well and promptly.

Below each of the descriptions, spaces permitted writing in three boys' names and three girls' names, members of the class who were considered "best fits" for each description. The children were not asked to write their nominations in order of merit. Children were instructed that any one child might be nominated for more than one characteristic. Independently of the children, the room teachers also filled in a nomination form. The 89-item scale of attitudes, now called the "Social Attitudes Scale," was administered to the same school classes a day or so following the "pupil nomination" form.

The children, who had spent all their school time with their principal teachers for the previous six months, were as well acquainted with each other and with their teachers as is generally found in public schools. They represented the socioeconomic distribution of the city. Two hundred and twenty-seven boys and 228 girls completed all of the schedules. Fortunately there were few absentees or incomplete records, and none among the crucial cases selected by the nomination form.

Although children were not required to nominate three of each sex, most of them supplied six names for each characteristic. A tally sheet was prepared for each of the four characteristics within each class group. The number of times a given child was mentioned for one characteristic constituted his score on that trait. Thus, the score did not distinguish whether a child's name appeared in first or

[1] This investigation was supported in part by a research grant (M-690) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, John E. Anderson and Dale B. Harris, Principal Investigators, and in part from funds supplied by the Institute of Child Welfare, University of Minnesota. The writer wishes to acknowledge the assistance of Dr. John E. Anderson and Dr. Harry Beilin in evaluating the theoretical relevance of items for the scale.

third place on a particular blank. Only a very few children received no mentions at all in their groups and, as is usually the case with Guess Who devices, five or six children in a group received the majority of the votes cast.

Rank-order correlations among the several traits for six class groups, one from each of six schools, show that the same names tended to appear high on the list for each of the four traits. For particular pairs of traits the medians of coefficients calculated from six school samples varied from +.78 to +.82[2], showing the order of relationships among the four traits to be much the same regardless of the pair of traits selected. Moreover, there was relatively little variation among the several school classes in the size of obtained correlations; values ranged typically from +.70 to +.89. The intercorrelations of scores based on one trait with scores obtained by summing scores for the other three characteristics ranged from +.78 to +.95 in various classes of 33 to 37 children.[3] Thus, no one characteristic appeared to be more closely related than another to the sum of scores on the remaining three aspects of responsibility. A total score obtained by summing the four trait scores appeared justified by this marked commonality among the aspects of responsibility.

To provide a criterion for selecting the attitude items associated with responsibility, 25 boys and 25 girls scoring highest on the sum of the four characteristics were selected from the total population. In every case children were selected who were also named for some component of responsibility on the teacher's ballot. In only three instances it was not possible to include the highest scoring child by the peer ballot when the teacher's judgment was added. These 50 responsible children were selected, two to six from each class, so as to include about a 10 per cent sample of each class enrollment. Boys and girls were selected in proportion to their distribution in the class. In the largest class groups an extra boy or girl was taken, since somewhat more than a 10 per cent sample was needed to produce 50 cases.

From each class an equal number of boys and girls were selected from the unmentioned or lowest scoring children, to constitute a criterion sample of "less responsible" children.[4]

DISCUSSION

The attitude tests for these criterion groups were segregated and analyzed. Proportions of each group subscribing to each item were read onto Lawshe's nomograph (6) and the discriminatory power of the item recorded. Fifty items satisfied a Lawshe value of +0.4 or better and have been retained. This represents approximately the 10 per cent level of significance. Of the 50 items, 21 appeared in the "How I Feel About Citizenship" scale used in the previous studies, and 29 were new items. Twenty-two of the items in the previous blank were not used in this schedule, either because they did not in our opinion have construct validity or did not hold up in the revalidation procedure described in this paper. This new blank, entitled "Social Attitudes Scale," was used in the 1954 Nobles County studies.[5] Retest reliabilities after an interval of four months ranged from +.60 to +.70 for various groups of eighth-and tenth-grade children. The items are reproduced in Table 1.

In another study, Rosemary Conzemius (1) used the "Social Attitudes Scale" with all 10-year-olds and high school seniors enrolled in the rural schools of another Minnesota county. She divided the children into those who belonged to 4-H Clubs, and those who did not. Table 2 compares data from her study with our results. Her non-4-H members tend to be slightly above our Nobles County rural youth, and her 4-H Club members appreciably exceed the Nobles County youth in comparable age groups. Conzemius found that 4-H Club members significantly exceeded the nonmem-

[2] Tables giving these data in full have been deposited with the American Documentation Institute. Order Document No. 5323 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 by 8 inch photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] *Ibid.*

[4] This method does not select *irresponsible* children by nomination for specifically described characteristics. Direct peer nominations for socially unacceptable behaviors sometimes have unfortunate social and public relations consequences. It is assumed that most such children would be included in the low count or unmentioned children when the positive nomination is requested.

[5] A county-wide survey of adjustment and mental health in children was conducted through funds supplied by the National Institute of Mental Health and the Institute of Child Welfare.

TABLE 1

ITEMS RETAINED IN THE "SOCIAL ATTITUDES SCALE"

1. It is always very important to finish anything one has started.
2. At school, it is easy to find things to do when the teacher doesn't give us enough work.
3. Police cars should be especially marked so that you can always see them coming.
4. It is no use worrying about current events or public affairs; I cannot do anything about them anyhow.
5. We ought to worry about our own country and let the rest of the world take care of itself.
*6. In school my behavior gets me into trouble.
*7. I am hardly ever on time for meals.
*8. I have been in trouble with the law or police.
9. When a person does not tell all his income in order to get out of paying some of his taxes, it is just as bad as stealing money from the government.
10. A person who does not vote when he can, is not a good citizen.
*11. I hardly ever get my school work done on time.
*12. I have played hookey from school.
13. Every citizen should take the time to find out about current events if it means giving up some spare time.
*14. In school I am sometimes sent to the principal for being bad.
15. Maybe some minority groups (Negroes, Indians, Mexicans, Jews, etc.) do get bad treatment, but it's no business of mine.
16. We ought to let Europe get out of its own mess.
17. People criticize me for wasting time.
*18. When I work on a committee, I usually let other people do most of the planning.
*19. I am often late for school.
20. If it is worth starting at all, it is worth finishing.
*21. I am the kind of person that people can count on.
*22. In school, I am one of those who can go on working even though the teacher is out of the room.
23. People can count on me to get things done, without checking on me.
*24. I am frequently chosen as a room helper or to run errands.
*25. I do my chores the very best I know how.
*26. I have been elected leader or president of my class.
27. Nothing is more important than to be honest with other people.
*28. My teacher often complains because I don't finish my work.
29. When you can't do a job, it is no use to try to find someone else to do it.
30. It is more important to get the job done than worry about hurting other people's feelings.
31. Why bother to vote when you can do so little with just your one vote.
32. "Never give a sucker an even break."
33. Letting your friends down is not so bad because you can't do good all the time for everybody.
34. Our country would be a lot better off if we didn't have elections and people didn't have to vote.
35. It's a good thing the Atlantic Ocean separates us from Europe because then we don't have to worry about them.
36. It's more important to work for the good of the team than to work for your own good.
*37. I would never let a friend down when he expects something of me.
38. People would be a lot better off if they could live far away from other people and never have to do anything for them.
39. Every person should give some of his time for the good of his town or city.
40. If everyone pitches in to do a job, it can always get done.
41. It is a good rule to do something for your neighbor if he does something for you.
42. Doing things which are important should come before things you enjoy doing.
43. When a person doesn't like something he is supposed to do he will try to get someone else to do it if he is smart.
44. Cheating on examinations is not so bad as long as nobody ever knows.
45. People have a real duty to take care of their parents when they are old, even if it costs a lot.
*46. I usually work things out for myself rather than get someone to show me how.
*47. I usually volunteer for special projects at school.
48. Children often get punished when they don't deserve it.
*49. When given a task I stick to it even if things I like to do better come along.
50. It doesn't really matter whether parents attend Parent-Teacher meetings regularly.

* Designates "personal reference" item.

TABLE 2

RESPONSIBILITY TEST SCORES FOR SEVERAL RURAL SAMPLES

Sample	Boys			Girls		
	N	$\bar{X}$	*SD*	*N*	$\bar{X}$	SD
10-year-olds, Nobles County	225	34.4	6.48	204	36.5	5.54
Conzemius' 10-year-olds						
Non-4-H members	85	35.8	4.51	92	38.5	5.48
4-H members	54	39.3	4.67	50	40.4	5.34
17-year-olds, Nobles County	119	37.8	5.96	112	41.6	4.43
Conzemius' high school seniors						
Non-4-H members	27	38.9	4.32	31	42.7	4.11
4-H members	34	41.1	4.67	29	43.6	3.19

bers in every comparison, though variances were not homogeneous in every instance. Whether this difference represents selection or the special effect of 4-H experience, she does not choose to say. Conzemius' experience is consistent with ours in finding greater sex differences than in our previous study, and a somewhat greater relationship of score to age than obtained in our previous work (3).

The "Social Attitudes Scale" yields a number of correlations with other measures which are of interest. Table 3 reports a number of these and shows that correlations of score with socio-economic status are negligible, but with measures of adjustment they are moderately high.

The correlation of the new scale with the old scale on two different samples of children who took the two schedules four years apart gives values of +.14 and +.32.

Whereas the "How I Feel About Citizenship" measure (1950) showed no age trend, a noticeable age relationship appears in the "Social Attitudes" test data (1954). From the youngest to the oldest age groups, the shift in mean score is approxmately one standard deviation of the total distribution. Furthermore, the relationship to age is more pronounced in girls than in boys. The differences between the means of the sexes is about one-fourth a standard deviation of the combined distributions in the younger ages and more than one-half a standard deviation in the oldest age groups. In the sense of this test, girls are perhaps better "socialized" than boys.

It will be recalled that less than half the items of the 1954 schedule appeared in the 1950 measure. It may be that the 1950 measure, based strictly on empirically selected items, contained a large random component, and it is true that the reliability coefficients were not large (of the order of +.64 to +.88 for a two-week interval). The 1954 measure, designed for construct as well as concurrent validity, seems to have included more age-related items. The reliability coefficients were comparable, being of the order of +.60 to +.70 for a four-month retest interval. Both measures, however, showed correlations with adjustment measures of the order of +.30 to +.40 and better, and the "Social Attitudes Scale" consistently yields the larger values. Furthermore, there is a suggestion that the Social Attitudes test tends also in the adolescent period towards an increased association with other measures, especially measures of adjustment. There may be more organization or responsibility as a focal characteristic as the child matures in his experience.

In 1950, 26 of the 43 items (60%) were personal reference items, couched in the first person and having to do with the child's own conduct. In 1954, only 18 of the 50 items (36%) were of this character. The tests for a stratified sample[6] in age groups 10, 12, 14, and 16 were recorded to provide separate scores on the 18 personal reference items (see marked items in Table 1) and on the 32 nonpersonal reference items (unmarked items in Table 1.) Table 4 makes it clear that the age increment in score on the Social Attitudes measure as a whole is due to age change in the nonpersonal reference items.

An analysis of another aspect of the Nobles County work[7] has shown that adjustment items of the psychoneurotic type are almost

TABLE 3

CORRELATIONS OF RESPONSIBILITY TEST SCORES WITH CERTAIN OTHER VARIABLES: BOYS AND GIRLS COMBINED

Variable	Ages 9 & 10	Ages 11 & 12	Ages 13 & 14	Ages 15 & 16
Father's education	.15	.17	.24	.15
Father's occupation (1954)	.06	.07	.11	.06
"How I Feel About Things"[a]				
Adjustment score	.38	.44	.42	.49
Adjustment-maturity score	.34	.45	.46	.47
Sentence Completion Test score[b]	.03	.16	.23	.27

[a] A schedule in the Nobles County study of attitude and interest items selected empirically to discriminate between well-adjusted and poorly adjusted children.

[b] This is also a Nobles County study instrument. It is scored according to the number of items showing positive affect in the completion supplied. This score has a low positive correlation with a criterion of adjustment.

TABLE 4

MEANS AND STANDARD DEVIATIONS FOR TOTAL SCORE ON THE SOCIAL ATTITUDES MEASURE AND FOR PERSONAL REFERENCE ITEMS AND NONPERSONAL REFERENCE ITEMS

Age	Sex	N	Total Score		Personal Reference Items		Nonpersonal Reference Items	
			Mean	SD	Mean	SD	Mean	SD
10	B	93	34.61	5.02	11.98	2.60	22.63	4.30
	G	91	37.08	4.75	13.63	1.82	23.45	3.90
12	B	95	36.94	5.14	12.49	2.11	24.44	4.06
	G	98	39.42	3.99	13.86	2.23	25.57	4.44
14	B	92	36.48	6.77	11.36	2.73	25.12	5.03
	G	93	40.53	4.48	13.39	2.07	27.14	3.39
16	B	94	37.74	6.56	12.01	2.59	25.73	4.85
	G	91	41.57	3.97	13.43	2.01	28.14	3.01

[6] For other research purposes a "standard sample" of 50 boys and 50 girls had been selected at each year of age from our larger supply of children (running about 300 in each age group). This sample was stratified according to occupation of the father, to represent the U.S. 1950 occupational distribution.

[7] As yet unpublished.

entirely of this personal reference type and are also free from an age relationship. Perhaps attitudes toward the community, society, and human relationships in general are modified by education, experience and maturity, while self-attitudes and personal habits are less amenable to change. Or, possibly, children progressively acquire more socially acceptable verbal statements about society and social relations as they move through school but fail to modify expressions about themselves, their self-feelings, and their own habits. Few studies of the "self image" have drawn this contrast, and this should prove a fruitful hypothesis for further work.

Correlations between the personal and nonpersonal subscores are modest, ranging from +.29 to +.53, and being characteristically around +.40.[5] There is no appreciable tendency for correlational values to change systematically with age, so that no argument can be offered concerning "integration" or "differentiation" of attitudes measured by these scales, over the age range represented in this study. Some information on the reliability of these subscales may be of interest. For a sample of 50 twelve-year-old girls the estimate of reliability obtained by Hoyt's analysis of variance method (5) was +.73. For the subscale of 18 personal reference items, the estimate of reliability was +.47, and for the 32 nonpersonal reference items the value was +.63.

SUMMARY

A scale of social attitudes was designed to discriminate children who have, with their peers, a reputation for responsibility as contrasted with children who have little reputation for responsibility. The resulting scale yields a measure which is substantially correlated with other measures of personal and social adjustment. Unlike an earlier scale designed for a similar purpose, the present scale shows a positive trend in mean score with age. Analysis revealed that this age trend results from a larger proportion of items in the present scale which refers to social and societal obligations (or at least do not contain a highly personal reference as indicated by the first personal singular pronoun). Items which have a highly personal reference, referring to self-feelings, attitudes, and reports of personal behavior, do not show an age trend. The two sets of items are modestly intercorrelated, on the order of +.40.

[5] American Documentation Institute Document No. 5323, *op. cit.*

REFERENCES

1. CONZEMIUS, ROSEMARY. A comparison of the development of social responsibility in rural boys and girls. Unpublished master's paper, School of Social Work, Univer. of Minnesota, June, 1955.
2. GOUGH, H. G., MCCLOSKEY, H., & MEEHL, P. E. A personality scale for social responsibility. *J. abnorm. soc. Psychol.*, 1952, **47**, 73–80.
3. HARRIS, D. B., CLARK, K. E., ROSE, A. M., & VALASEK, FRANCES. The measurement of responsibility in children. *Child Develpm.*, 1954, **25**, 21–28.
4. HARRIS, D. B., CLARK, K. E., ROSE, A. M., & VALASEK, FRANCES. The relationship of children's home duties to an attitude of responsibility. *Child Develpm.*, 1954, **25**, 29–33.
5. HOYT, C. J. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **3**, 153–60.
6. LAWSHE, C. H. A nomograph for estimating the validity of test items. *J. appl. Psychol.*, 1942, **26**, 846–849.

Received January 21, 1957.

A CROSS-CULTURAL SURVEY OF SOME SEX DIFFERENCES IN SOCIALIZATION[1]

HERBERT BARRY III, MARGARET K. BACON, AND IRVIN L. CHILD

Yale University

IN OUR society, certain differences may be observed between the typical personality characteristics of the two sexes. These sex differences in personality are generally believed to result in part from differences in the way boys and girls are reared. To the extent that personality differences between the sexes are thus of cultural rather than biological origin, they seem potentially susceptible to change. But how readily susceptible to change? In the differential rearing of the sexes does our society make an arbitrary imposition on an infinitely plastic biological base, or is this cultural imposition found uniformly in all societies as an adjustment to the real biological differences between the sexes? This paper reports one attempt to deal with this problem.

DATA AND PROCEDURES

The data used were enthnographic reports, available in the anthropological literature, about socialization practices of various cultures. One hundred and ten cultures, mostly nonliterate, were studied. They were selected primarily in terms of the existence of adequate ethnographic reports of socialization practices and secondarily so as to obtain a wide and reasonably balanced geographical distribution. Various aspects of socialization of infants and children were rated on a 7-point scale by two judges (Mrs. Bacon and Mr. Barry). Where the ethnographic reports permitted, separate ratings were made for the socialization of boys and girls. Each rating was indicated as either confident or doubtful; with still greater uncertainty, or with complete lack of evidence, the particular rating was of course not made at all. We shall restrict the report of sex difference ratings to cases in which both judges made a confident rating. Also omitted is the one instance where the two judges reported a sex difference in opposite directions, as it demonstrates only unreliability of judgment. The number of cultures that meet these criteria is much smaller than the total of 110; for the several variables to be considered, the number varies from 31 to 84.

The aspects of socialization on which ratings were made included:

1. Several criteria of attention and indulgence toward infants.

2. Strength of socialization from age 4 or 5 years until shortly before puberty, with respect to five systems of behavior; strength of socialization was defined as the combination of positive pressure (rewards for the behavior) plus negative pressure (punishments for lack of the behavior). The variables were:

(*a*) Responsibility or dutifulness training. (The data were such that training in the performance of chores in the productive or domestic economy was necessarily the principal source of information here; however, training in the performance of other duties was also taken into account when information was available.)

(*b*) Nurturance training, i.e., training the child to be nurturant or helpful toward younger siblings and other dependent people.

(*c*) Obedience training.

(*d*) Self-reliance training.

(*e*) Achievement training, i.e., training the child to orient his behavior toward standards of excellence in performance, and to seek to achieve as excellent a performance as possible.

Where the term "no sex difference" is used here, it may mean any of three things: (*a*) the judge found separate evidence about the training of boys and girls on this particular variable, and judged it to be identical; (*b*) the judge found a difference between the training of boys and girls, but not great enough for the sexes to be rated a whole point apart on a 7-point scale; (*c*) the judge found evidence only about the training of "children" on this variable, the ethnographer not reporting separately about boys and girls.

SEX DIFFERENCES IN SOCIALIZATION

On the various aspects of attention and indulgence toward infants, the judges almost always agreed in finding no sex difference. Out of 96 cultures for which the ratings included the infancy period, 88 (92%) were rated with no sex difference by either judge for any of those variables. This result is consistent with the

[1] This research is part of a project for which financial support was provided by the Social Science Research Council and the Ford Foundation. We are greatly indebted to G. P. Murdock for supplying us with certain data, as indicated below, and to him and Thomas W. Maretzki for suggestions that have been used in this paper.

TABLE 1

RATINGS OF CULTURES FOR SEX DIFFERENCES ON FIVE VARIABLES OF CHILDHOOD SOCIALIZATION PRESSURE

Variable	Number of Cultures	Both Judges Agree in Rating the Variable Higher in		One Judge Rates No Difference, One Rates the Variable Higher in		Percentage of Cultures with Evidence of Sex Difference in Direction of		
		Girls	Boys	Girls	Boys	Girls	Boys	Neither
Nurturance	33	17	0	10	0	82%	0%	18%
Obedience	69	6	0	18	2	35%	3%	62%
Responsibility	84	25	2	26	7	61%	11%	28%
Achievement	31	0	17	1	10	3%	87%	10%
Self-reliance	82	0	64	0	6	0%	85%	15%

point sometimes made by anthropologists that "baby" generally is a single status undifferentiated by sex, even though "boy" and "girl" are distinct statuses.

On the variables of childhood socialization, on the other hand, a rating of no sex difference by both judges was much less common. This finding of no sex difference varied in frequency from 10% of the cultures for the achievement variable up to 62% of the cultures for the obedience variable, as shown in the last column of Table 1. Where a sex difference is reported, by either one or both judges, the difference tends strongly to be in a particular direction, as shown in the earlier columns of the same table. Pressure toward nurturance, obedience, and responsibility is most often stronger for girls, whereas pressure toward achievement and self-reliance is most often stronger for boys.

For nurturance and for self-reliance, all the sex differences are in the same direction. For achievement there is only one exception to the usual direction of difference, and for obedience only two; but for responsibility there are nine. What do these exceptions mean? We have reexamined all these cases. In most of them, only one judge had rated the sexes as differently treated (sometimes one judge, sometimes the other), and in the majority of these cases both judges were now inclined to agree that there was no convincing evidence of a real difference. There were exceptions, however, especially in cases where a more formal or systematic training of boys seemed to imply greater pressure on them toward responsibility. The most convincing cases were the Masai and Swazi, where both judges had originally agreed in rating responsibility pressures greater in boys than in girls. In comparing the five aspects of socialization we may conclude that responsibility shows by far the strongest evidence of real variation in the direction of sex difference, and obedience much the most frequently shows evidence of no sex difference at all.

In subsequent discussion we shall be assuming that the obtained sex differences in the socialization ratings reflect true sex differences in the cultural practices. We should consider here two other possible sources of these rated differences.

1. The ethnographers could have been biased in favor of seeing the same pattern of sex differences as in our culture. However, most anthropologists readily perceive and eagerly report novel and startling cultural features, so we may expect them to have reported unusual sex differences where they existed. The distinction between matrilineal and patrilineal, and between matrilocal and patrilocal cultures, given prominence in many ethnographic reports, shows an awareness of possible variations in the significance of sex differences from culture to culture.

2. The two judges could have expected to find in other cultures the sex roles which are familiar in our culture and inferred them from the material on the cultures. However, we have reported only confident ratings, and such a bias seems less likely here than for doubtful ratings. It might be argued, moreover, that bias has more opportunity in the cases ambiguous enough so that only one judge reported a sex difference, and less opportunity in the cases where the evidence is so clear that both judges agree. Yet in general, as may be seen in Table 1, the deviant cases are somewhat more frequent among the cultures where only one judge reported a sex difference.

The observed differences in the socialization of boys and girls are consistent with certain universal tendencies in the differentiation of adult sex role. In the economic sphere, men are more frequently allotted tasks that involve leaving home and engaging in activities where

a high level of skill yields important returns; hunting is a prime example. Emphasis on training in self-reliance and achievement for boys would function as preparation for such an economic role. Women, on the other hand, are more frequently allotted tasks at or near home that minister most immediately to the needs of others (such as cooking and water carrying); these activities have a nurturant character, and in their pursuit a responsible carrying out of established routines is likely to be more important than the development of an especially high order of skill. Thus training in nurturance, responsibility, and, less clearly, obedience, may contribute to preparation for this economic role. These consistencies with adult role go beyond the economic sphere, of course. Participation in warfare, as a male prerogative, calls for self-reliance and a high order of skill where survival or death is the immediate issue. The childbearing which is biologically assigned to women, and the child care which is socially assigned primarily to them, lead to nurturant behavior and often call for a more continuous responsibility than do the tasks carried out by men. Most of these distinctions in adult role are not inevitable, but the biological differences between the sexes strongly predispose the distinction of role, if made, to be in a uniform direction.[2]

The relevant biological sex differences are conspicuous in adulthood but generally not in childhood. If each generation were left entirely to its own devices, therefore, without even an older generation to copy, sex differences in role would presumably be almost absent in childhood and would have to be developed after puberty at the expense of considerable relearning on the part of one or both sexes. Hence, a pattern of child training which foreshadows adult differences can serve the useful function of minimizing what Benedict termed "discontinuities in cultural conditioning" (1).

The differences in socialization between the sexes in our society, then, are no arbitrary custom of our society, but a very widespread adaptation of culture to the biological substratum of human life.

[2] For data and interpretations supporting various arguments of this paragraph, see Mead (2), Murdock (3), and Scheinfeld (6).

VARIATIONS IN DEGREE OF SEX DIFFERENTIATION

While demonstrating near-universal tendencies in direction of difference between the socialization of boys and girls, our data do not show perfect uniformity. A study of the variations in our data may allow us to see some of the conditions which are associated with, and perhaps give rise to, a greater or smaller degree of this difference. For this purpose, we classified cultures as having relatively large or small sex difference by two different methods, one more inclusive and the other more selective. In both methods the ratings were at first considered separately for each of the five variables. A sex difference rating was made only if both judges made a rating on this variable and at least one judge's rating was confident.

In the more inclusive method the ratings were dichotomized, separately for each variable, as close as possible to the median into those showing a large and those showing a small sex difference. Thus, for each society a large or a small sex difference was recorded for each of the five variables on which a sex difference rating was available. A society was given an over-all classification of large or small sex difference if it had a sex difference rating on at least three variables and if a majority of these ratings agreed in being large, or agreed in being small. This method permitted classification of a large number of cultures, but the grounds for classification were capricious in many cases, as a difference of only one point in the rating of a single variable might change the over-all classification of sex difference for a culture from large to small.

In the more selective method, we again began by dichotomizing each variable as close as possible to the median; but a society was now classified as having a large or small sex difference on the variable only if it was at least one step away from the scores immediately adjacent to the median. Thus only the more decisive ratings of sex difference were used. A culture was classified as having an over-all large or small sex difference only if it was given a sex difference rating which met this criterion on at least two variables, and only if all such ratings agreed in being large, or agreed in being small.

We then tested the relation of each of these dichotomies to 24 aspects of culture on which

TABLE 2
CULTURE VARIABLES CORRELATED WITH LARGE SEX DIFFERENCE IN SOCIALIZATION, SEPARATELY FOR TWO TYPES OF SAMPLE

Variable	More Selective Sample		More Inclusive Sample	
	ϕ	N	ϕ	N
Large animals are hunted	.48*	(34)	.28*	(72)
Grain rather than root crops are grown	.82**	(20)	.62**	(43)
Large or milking animals rather than small animals are kept	.65*	(19)	.43*	(35)
Fishing unimportant or absent	.42*	(31)	.19	(69)
Nomadic rather than sedentary residence	.61**	(34)	.15	(71)
Polygyny rather than monogamy	.51*	(28)	.38**	(64)

* $p < .05$
** $p < .01$.

Note.—The variables have been so phrased that all correlations are positive. The phi coefficient is shown, and in parentheses, the number of cases on which the comparison was based. Significance level was determined by χ^2, or Fisher's exact test where applicable, using in all cases a two-tailed test.

Murdock has categorized the customs of most of these societies[3] and which seemed of possible significance for sex differentiation. The aspects of culture covered include type of economy, residence pattern, marriage and incest rules, political integration, and social organization. For each aspect of culture, we grouped Murdock's categories to make a dichotomous contrast (sometimes omitting certain categories as irrelevant to the contrast). In the case of some aspects of culture, two or more separate contrasts were made (e.g., under form of marriage we contrasted monogamy with polygyny, and also contrasted sororal with nonsororal polygyny). For each of 40 comparisons thus formed, we prepared a 2 x 2 frequency table to determine relation to each of our sex-difference dichotomies. A significant relation was found for six of these 40 aspects of culture with the more selective dichotomization of over-all sex difference. In four of these comparisons, the relation to the more inclusive dichotomization was also significant. These relationships are all given in Table 2, in the form of phi coefficients, along with the outcome of testing significance by the use of χ^2 or Fisher's exact test. In trying to interpret these findings, we have also considered the nonsignificant correlations with other variables, looking for consistency and inconsistency with the general implications of the significant findings. We have arrived at the following formulation of results:

1. Large sex difference in socialization is associated with an economy that places a high premium on the superior strength, and superior development of motor skills requiring strength, which characterize the male. Four of the correlations reported in Table 2 clearly point to this generalization: the correlations of large sex difference with the hunting of large animals, with grain rather than root crops, with the keeping of large rather than small domestic animals, and with nomadic rather than sedentary residence. The correlation with the unimportance of fishing may also be consistent with this generalization, but the argument is not clear.[4] Other correlations consistent with the generalization, though not statistically significant, are with large game hunting rather than gathering, with the hunting of large game rather than small game, and with the general importance of all hunting and gathering.

2. Large sex difference in socialization appears to be correlated with customs that make for a large family group with high cooperative interaction. The only statistically significant correlation relevant here is that with polygyny rather than monogamy. This generalization is, however, supported by several substantial correlations that fall only a little short of being statistically significant. One of these is a

[3] These data were supplied to us directly by Professor Murdock.

[4] Looking (with the more inclusive sample) into the possibility that this correlation might result from the correlation between fishing and sedentary residence, a complicated interaction between these variables was found. The correlation of sex differentiation with absence of fishing is found only in nomadic societies, where fishing is likely to involve cooperative activity of the two sexes, and its absence is likely to mean dependence upon the male for large game hunting or herding large animals (whereas in sedentary societies the alternatives to fishing do not so uniformly require special emphasis on male strength). The correlation of sex differentiation with nomadism is found only in nonfishing societies; here nomadism is likely to imply large game hunting or herding large animals, whereas in fishing societies nomadism evidently implies no such special dependence upon male strength. Maximum sex differentiation is found in nomadic nonfishing societies (15 with large difference and only 2 with small) and minimum sex differentiation in nomadic fishing societies (2 with large difference and 7 with small difference). These findings further strengthen the argument for a conspicuous influence of the economy upon sex differentiation.

correlation with sororal rather than nonsororal polygyny; Murdock and Whiting (4) have presented indirect evidence that co-wives generally show smoother cooperative interaction if they are sisters. Correlations are also found with the presence of either an extended or a polygynous family rather than the nuclear family only; with the presence of an extended family; and with the extreme contrast between maximal extension and no extension of the family. The generalization is also to some extent supported by small correlations with wide extension of incest taboos, if we may presume that an incest taboo makes for effective unthreatening cooperation within the extended family. The only possible exception to this generalization, among substantial correlations, is a near-significant correlation with an extended or polygynous family's occupying a cluster of dwellings rather than a single dwelling.[5]

In seeking to understand this second generalization, we feel that the degree of social isolation of the nuclear family may perhaps be the crucial underlying variable. To the extent that the nuclear family must stand alone, the man must be prepared to take the woman's role when she is absent or incapacitated, and vice versa. Thus the sex differentiation cannot afford to be too great. But to the extent that the nuclear family is steadily interdependent with other nuclear families, the female role in the household economy can be temporarily taken over by another woman, or the male role by another man, so that sharp differentiation of sex role is no handicap.

The first generalization, which concerns the economy, cannot be viewed as dealing with material completely independent of the ratings of socialization. The training of children in their economic role was often an important part of the data used in rating socialization variables, and would naturally vary according to the general economy of the society. We would stress, however, that we were by no means using the identical data on the two sides of our comparison; we were on the one hand judging data on the socialization of children and on the other hand using Murdock's judgments on the economy of the adult culture. In the case of the second generalization, it seems to us that there was little opportunity for information on family and social structure to have influenced the judges in making the socialization ratings.

Both of these generalizations contribute to understanding the social background of the relatively small difference in socialization of boys and girls which we believe characterizes our society at the present time. Our mechanized economy is perhaps less dependent than any previous economy upon the superior average strength of the male. The nuclear family in our society is often so isolated that husband and wife must each be prepared at times to take over or help in the household tasks normally assigned to the other. It is also significant that the conditions favoring low sex differentiation appear to be more characteristic of the upper segments of our society, in socioeconomic and educational status, than of lower segments. This observation may be relevant to the tendency toward smaller sex differences in personality in higher status groups (cf. Terman and Miles, 8).

The increase in our society of conditions favoring small sex difference has led some people to advocate a virtual elimination of sex differences in socialization. This course seems likely to be dysfunctional even in our society. Parsons, Bales, *et al.* (5) argue that a differentiation of role similar to the universal pattern of sex difference is an important and perhaps inevitable development in any social group, such as the nuclear family. If we add to their argument the point that biological differences between the sexes make most appropriate the usual division of those roles between the sexes, we have compelling reasons to expect that the decrease in differentiation of adult sex role will not continue to the vanishing point. In our training of children, there may now be less differentiation in sex role than characterizes

[5] We think the reverse of this correlation would be more consistent with our generalization here. But perhaps it may reasonably be argued that the various nuclear families composing an extended or polygynous family are less likely to develop antagonisms which hinder cooperation if they are able to maintain some physical separation. On the other hand, this variable may be more relevant to the first generalization than to the second. Occupation of a cluster of dwellings is highly correlated with presence of herding and with herding of large rather than small animals, and these economic variables in turn are correlated with large sex difference in socialization. Occupation of a cluster of dwellings is also correlated with polygyny rather than monogamy and shows no correlation with sororal vs. nonsororal polygyny.

adult life—so little, indeed, as to provide inadequate preparation for adulthood. This state of affairs is likely to be especially true of formal education, which is more subject to conscious influence by an ideology than is informal socialization at home. With child training being more oriented toward the male than the female role in adulthood, many of the adjustment problems of women in our society today may be partly traced to conflicts growing out of inadequate childhood preparation for their adult role. This argument is nicely supported in extreme form by Spiro's analysis of sex roles in an Israeli kibbutz (7). The ideology of the founders of the kibbutz included the objective of greatly reducing differences in sex role. But the economy of the kibbutz is a largely nonmechanized one in which the superior average strength of men is badly needed in many jobs. The result is that, despite the ideology and many attempts to implement it, women continue to be assigned primarily to traditional "women's work," and the incompatibility between upbringing or ideology and adult role is an important source of conflict for women.

Note on regional distribution. There is marked variation among regions of the world in typical size of sex difference in socialization. In our sample, societies in North America and Africa tend to have large sex difference, and societies in Oceania to have small sex difference. Less confidently, because of the smaller number of cases, we can report a tendency toward small sex differences in Asia and South America as well. Since most of the variables with which we find the sex difference to be significantly correlated have a similar regional distribution, the question arises whether the correlations might better be ascribed to some quite different source having to do with large regional similarities, rather than to the functional dependence we have suggested. As a partial check, we have tried to determine whether the correlations we report in Table 2 tend also to be found strictly within regions. For each of the three regions for which we have sizable samples (North America, Africa, and Oceania) we have separately plotted 2 x 2 tables corresponding to each of the 6 relationships reported in Table 2. (We did this only for the more inclusive sample, since for the more selective sample the number of cases within a region would have been extremely small.) Out of the 18 correlations thus determined, 11 are positive and only 3 are negative (the other 4 being exactly zero). This result clearly suggests a general tendency for these correlations to hold true within regions as well as between regions, and may lend further support to our functional interpretation.

SUMMARY

A survey of certain aspects of socialization in 110 cultures shows that differentiation of the sexes is unimportant in infancy, but that in childhood there is, as in our society, a widespread pattern of greater pressure toward nurturance, obedience, and responsibility in girls, and toward self-reliance and achievement striving in boys. There are a few reversals of sex difference, and many instances of no detectable sex difference; these facts tend to confirm the cultural rather than directly biological nature of the differences. Cultures vary in the degree to which these differentiations are made; correlational analysis suggests some of the social conditions influencing these variations, and helps in understanding why our society has relatively small sex differentiation.

REFERENCES

1. BENEDICT, RUTH. Continuities and discontinuities in cultural conditioning. *Psychiatry*, 1938, **1**, 161–167.
2. MEAD, MARGARET. *Male and female.* New York: Morrow, 1949.
3. MURDOCK, G. P. Comparative data on the division of labor by sex. *Social Forces*, 1937, **16**, 551–553.
4. MURDOCK, G. P., & WHITING, J. W. M. Cultural determination of parental attitudes: The relationship between the social structure, particularly family structure and parental behavior. In M. J. E. Senn (Ed.), *Problems of infancy and childhood: Transactions of the Fourth Conference,* March 6–7, 1950. New York: Josiah Macy, Jr. Foundation, 1951. Pp. 13–34.
5. PARSONS, T., BALES, R. F., *et al. Family, socialization and interaction process.* Glencoe, Ill.: Free Press, 1955.
6. SCHEINFELD, A. *Women and men.* New York: Harcourt, Brace, 1944.
7. SPIRO, M. E. *Kibbutz: Venture in Utopia.* Cambridge: Harvard Univer. Press, 1956.
8. TERMAN, L. M., & MILES, CATHERINE C. *Sex and personality.* New York: McGraw-Hill, 1936.

Received February 6, 1957.

AVOIDANCE LEARNING OF PERCEPTUAL DEFENSE AND VIGILANCE[1]

DON E. DULANY, JR.
University of Illinois

THOSE who argue for a perceptual defense mechanism analogous to repression and those who believe that the adaptive value of sensitization to threat suggests a vigilance mechanism can both find support in a wealth of controversial data (3, 7)—as well as can those who would deny both views (4, 11, 12, 13). Much of the confusion has resulted from the use of linguistic stimuli whose significance to the subject is a matter of dispute. So long as the history of the subjects' commerce with the stimulus materials remains uncertain, convenient assumptions permit any of the several views to be salvaged in the face of most data. The chief object of the present study is to produce experimentally the learning of perceptual defense and vigilance, and in so doing, to relate the mechanisms to antecedent experimental procedures.

Knowing that a stimulus is anxiety arousing does not in itself permit us to predict the perceptual reaction to that stimulus. Lazarus, Eriksen, and Fonda (6) have demonstrated both perceptual defense and vigilance reactions to the same hostile or sexual, and presumably threatening, material; which reaction occurs is predicted from diagnostic categories. Encountering stimuli with sexual or hostile meaning is no novelty to the adult subject, who has had ample opportunity for learning to handle any anxiety they engender. We might guess that one kind of subject has learned to defend against threatening stimuli because with his particular experiences and personal economy that reaction has somehow been to his advantage. Another has learned perceptual vigilance because that reaction served him well. The specific learning mechanism involved has not been made clear. Hence, a second purpose of this study is to offer an account in terms of behavior theory of how perceptual defense and vigilance may be learned and to test two hypotheses derived from that account.

This analysis extends to perception a two-stage conception of avoidance learning that is essentially in agreement with the views of Mowrer (9) and of Solomon (16). Once an anxiety response has been classically conditioned to a previously neutral stimulus, any number of perceptual reactions are theoretically possible—vigilance and defense, as well as many qualitative distortions. The convenience of certain psychophysical methods has led to great experimental interest in two such reactions along the single dimension of recognizability. We may think of these potential perceptual reactions to a threatening stimulus as ordered in a hierarchy of probability of occurrence. There is some evidence that perceptual vigilance holds the dominant position in an initial hierarchy of perceptual reactions to threat. Pustell (14) produced perceptual vigilance experimentally by classically associating an electric shock with a number of neutral geometric figures. Lazarus and McCleary (5) classically conditioned GSRs to nonsense syllables, obtaining their threshold measures under threat of further punishment, but failed to find a significant difference between thresholds for their shock syllables and a number of nonshock syllables. Could a reluctance to name the shock syllable have cancelled out a primitive vigilance reaction? Lysak (8) replicated their finding, then in another experiment removed the electrodes before obtaining threshold measures and found thresholds for his shock syllables to be lower than those for nonshock syllables. These findings suggest that perceptual vigilance is dominant in an initial hierarchy of perceptual reactions to a threatening stimulus, although response suppression may obscure the effect when there is a continuing imminent possibility of punishment. The superior adaptive value of such a mechanism is obvious.

Whatever perceptual reaction is initially dominant, it should be possible to realign this hierarchy by introducing selective reinforcement. The perceptual reaction that is instrumental to the avoidance of punishment and reduction of the conditioned anxiety should be

[1] This article is based on a doctoral dissertation submitted to the Department of Psychology, University of Michigan. I wish to express my gratitude to Dr. E. L. Walker for his generous advice and guidance as chairman of my doctoral committee.

strengthened. In the course of ordinary experience, perceptual defense and vigilance occur in the presence of varied and competing stimuli. They name two kinds of imbalance among a set of competing perceptual responses. If one perceptual response is punished while competing perceptual responses are instrumental to the avoidance of punishment and a reduction in anxiety, perceptual defense should be learned, since competing perceptual responses should tend to crowd out the critical perceptual response. On the other hand, perceptual vigilance should be learned when one perceptual response is instrumental to avoidance of punishment while competing perceptual responses are punished. Specifically, the following hypotheses are tested: (*a*) When one perceptual response is followed by punishment and competing perceptual responses are instrumental to avoidance of punishment, the punished response becomes weaker as compared with the competing perceptual responses; (*b*) When one perceptual response is instrumental to avoidance of punishment and competing perceptual responses are punished, the avoidance response becomes stronger as compared with the competing perceptual responses.

METHOD

Subjects

The *S*s were 32 undergraduates at the University of Michigan, evenly divided between the defense and vigilance training procedures. They were randomly assigned to the two procedures, the only requirement being that an equal number of men and women participate in each. All *S*s were volunteers, and none was aware of the purpose of the experiment.

Stimulus Materials

Four rather meaningless, emotionally neutral figures—a circle, diamond, square, and triangle—were cut from black Munsell paper and mounted on white cards. The figures were arranged top, bottom, left, and right, on each of four cards, so that each figure appeared in all four positions. Hence, if each of the four cards was presented to the *S* an equal number of times, each figure would appear an equal number of times in each of the four positions. In this way, positional (see 10) and frequency effects were controlled. Prior to this experiment, the figures had been roughly equated for discriminability by Thomas Pustell.

Apparatus

A Gerbrands tachistoscope, designed for individual use, was adjusted for a constant exposure interval of .12 sec. throughout all experimental series. With a pre-exposure field of 6.29 ft. lamberts, an exposure field of 0.425 ft. lamberts, and geometrical forms of less than one inch diameter, this exposure time produced a level of awareness too low for the figures to be readily named.

The electric shock applicator contained four pairs of electrodes spaced about an inch apart and was designed to be strapped to the leg of the *S* just above the ankle. In order to discourage adaptation and at the same time prevent burns, the shock could be sent through first one pair of electrodes and then another from trial to trial. The shock circuits and applicator were designed by Dr. Carl Brown.

Defense Training Experiment

Level of awareness series. The experimental session began for each *S* with an assessment of his level of awareness. A particular exposure duration, together with illumination level, does not define a single level of awareness for more than one *S*. To ascertain whether level of awareness was related to the amount and character of the experimental effect, the *S* was given a simple discrimination task. A forced-choice technique developed by Blackwell (1) and modified by Blum (2) was selected for its sensitivity in measuring discrimination without awareness. It has the additional value of permitting a number of stimuli to be presented at once, a condition that more truly represents most perceptual situations.

The *S* was informed that he would be shown a number of cards, but that they would be flashed too briefly for him to see much. On each trial, *E* asked *S* to "Find the square" or "Find the diamond," etc. The *S* responded with "Top," "Bottom," "Left," "Right," depending on where he thought he saw the figure. He was instructed on each trial to focus on the fixation point in the center of the field. Sixty-four trials were run, and *S*'s level of awareness was measured by his percentage of correct identifications.

Baseline series. Although the figures were roughly equal in recognizability for a group collectively, individuals find one figure or another more salient because of their peculiar perceptual histories and predilections. For this reason an empirical baseline of recognizability of each figure was obtained for every *S*. The procedure for this series was like that of the preceding in every way except that the *S* was instructed to "tell me on each trial which figure is *clearest*—which one is most *recognizable*." He was again asked to convey his choice by saying "Top," "Bottom," "Left," or "Right," and to follow his hunch whenever uncertain. Any experimental effect produced

by the following training procedure should register as a departure from this empirical baseline.

Training series. The defense training procedure was essentially like the baseline series, with the important difference that reinforcement was introduced. One of the geometrical figures, selected arbitrarily, was designated the critical stimulus. Each of the figures was the critical one for four of the 16 *S*s. As in the preceding series, *S* was instructed to report on each trial which of the four figures was most recognizable by indicating its position. When the critical stimulus was selected, an electric shock followed two seconds after exposure of the figures and lasted another two seconds. Selection of one of the other three figures as most recognizable resulted in avoidance of the shock.

Circuits were designed so as to permit *S*'s pressing of one of four buttons to reveal his choice of figures and control presentation of the shock. The critical stimulus, of course, varied in position from trial to trial, and therefore no single one of the four responses was always punished. The *E* noted the position of the critical stimulus before each trial, and adjusted a commutator, thereby determining which of the four buttons would activate the shock circuit.

For the purposes of this experiment it was imperative that the shock be punishing. It was necessary that it be decidedly unpleasant, even painful for best results, but short of incapacitating. Since individual pain thresholds for electric shock vary so widely, the intensity of shock was set at a subjective level for each *S*. After he had been assured that the shock was quite harmless and encouraged to take as much as he could stand, the shock was set at the level each *S* judged to be "just this side of beyond endurance." By phenomenological concensus of 32 *S*s, the shock was decidedly unpleasant, even painful, and like the jab of a hot needle.

Sixty-four trials were run, approximately one every 10 seconds. After each block of 16 trials, *S* rested briefly.

Assessment series. Immediately after the training series, *S* was allowed to remove the shock applicator and was assured that there would be no more shock. The instructions and procedure for the assessment series were identical with those for the baseline series. The same cards were presented for 64 trials, and on each trial *S* indicated which figure was most recognizable. Finally, he was questioned for any knowledge of the rule by which he had been shocked.

Vigilance Training Experiment

The vigilance training experiment repeated the defense training experiment with one important difference: the reinforcement prescription was reversed. When *S* selected one of the three noncritical stimuli as most recognizable, he was shocked; when he selected the critical stimulus, he avoided the shock.

Results and Discussion

Relative perceptual response strength for each figure was measured by the frequency with which that figure was judged clearer than competing stimuli. The *S*s of the defense training procedure should, if the first hypothesis is correct, select the critical figure less frequently in the assessment series, after training, than they did in the baseline series. Amount of shift in frequency of response ranged from 0 to −14 with a mean of −5.8. In order to evaluate the group shift, it was noted that 14 of the 16 *S*s shifted in the predicted direction while two remained unchanged. By the sign test this shift is significant with a $p < .01$.[2] These findings confirm Hypothesis I.

Hypothesis II leads to the contrary expectation that *S*s of the vigilance training procedure should select the critical stimulus more frequently during the assessment series than during the baseline series. Magnitude of shift ranged from −7 to +36 with a mean of +7.0. Thirteen of 16 *S*s shifted in the anticipated direction, and this shift is significant by the sign test with a $p < .025$. Thus the second hypothesis is supported.

We would also like to know what progressive changes take place throughout the training series, whether these changes follow a typical learning curve, and whether there is any sign of extinction in the assessment series. Consequently, pooled frequency of selection of the critical stimulus was graphed as a function of trials.

For the defense training group (see Fig. 1), frequency of selection drops precipitously from the baseline to the early part of the training series. Apparently the learning takes place very rapidly, although the abruptness of the drop

[2] Since the direction of experimental shift was predicted, one-tail p values are reported.

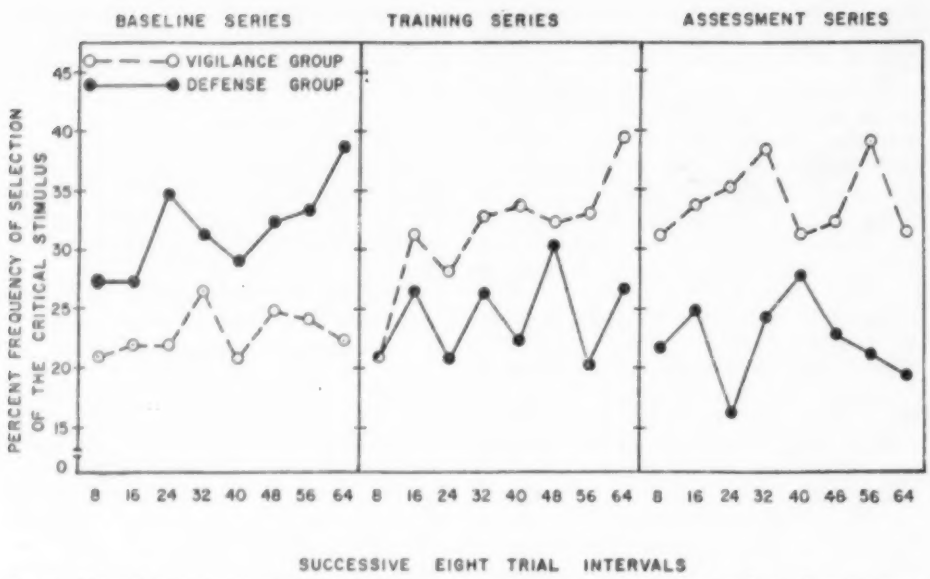

FIG. 1. MEAN PERFORMANCE OF THE DEFENSE AND VIGILANCE GROUPS THROUGHOUT EACH EXPERIMENTAL SERIES

may be partly due to a chance increase in frequency toward the end of the baseline series. In the vigilance training series, frequency of selection increases throughout the training series in a manner strongly suggestive of a learning curve. Since an anxiety reaction is very probably conditioned early in both training series, or even earlier, when the shock is adjusted to the appropriate subjective level, there is some reason to expect the instrumental learning of defense to proceed more rapidly than the learning of vigilance. Theoretically, vigilance is acquired as a single, critical, perceptual response, strengthened by anxiety reduction attendant upon shock avoidance; this is the kind of learning that ought to be describable by a familiar learning curve. The learning of perceptual defense, however, proceeds as the three competing perceptual responses are strengthened, by anxiety reduction, at the expense of the critical perceptual response. There is less basis for anticipating the function for this kind of learning. That the learning of defense did occur more rapidly than the learning of vigilance might be explained by the greater probability of reinforcement in the defense series. It is also likely that the greater number of shocks in the vigilance series produced a higher level of anxiety in the vigilance group. Conceivably, level of anxiety drive could be related to the rapidity and course of learning. Finally, for neither procedure is there any evidence of extinction in the assessment series; and in view of the well-documented resistance to extinction of avoidance responses (15, 17) this is not too surprising.

A problem of interpretation arises from the fact that the empirical baseline values differ somewhat from their expected value of 25%–31.6% for the defense group and 23.0% for the vigilance group. While there is no reasonable alternative but to attribute this disparity to sampling fluctuation, it is important to distinguish two possible sources of error. Have two deviant samples been drawn from a population of responses that mark the baseline? Or have such deviant samples been drawn from a population of *S*s varying in perceptual sensitivity to the critical stimuli? In the former case, we should expect some regression toward the value of 25% in the assessment series, quite apart from any effect of the training procedures. If the latter possibility is the case, then there is no problem, and the experimental effect is the more impressive for having overridden figure preferences. It is

possible, for purposes of analysis, to eliminate those four *Ss* with the highest frequencies at the defense baseline and the three *Ss* lowest at the vigilance baseline, leaving the two baseline means approximately equal. Of the 12 *Ss* remaining in the defense group, 10 selected the critical stimulus less frequently during the assessment series than at the baseline. By the sign test this shift is significant with $p < .025$. Ten of the 13 *Ss* remaining in the vigilance group shifted in the predicted direction ($p < .05$). It seems clear that the regression hypothesis cannot account for these results.

How justified is the interpretation that these training procedures have changed the strength of perceptual responses rather than *S*'s disposition to make a report that has invariably brought him a painful electric shock? If *S* at any time discerned a connection between pressing the button for a square and being shocked, it would be only prudent for him to take a good look on each trial, then press one of the other buttons. For a number of reasons this suggested process cannot account for these results. Only two of 32 *Ss* could name the critical figure or verbalize the rule by which they had been shocked. Typical of their replies when questioned were, "Maybe there was a pattern to it, but I don't know what it was" and "I didn't know what I did or saw would have anything to do with it." This is, in fact, learning without awareness. Moreover, if the *S* were deliberately falsifying his report, then during the assessment series, after the shock applicator had been removed, he should have abandoned this tactic; frequency of selection of the critical stimulus would have returned to what it was at the baseline. It is evident from the data that this did not happen. Furthermore, the greater his awareness, the more information he has to act on and the more successful he should be; yet no relation was found between level of awareness and magnitude of experimental effect.

If the theoretical assumptions of the study are correct, something like the following occurred. An anxiety reaction produced by the shock was conditioned to the stimulus figures. It seems likely that anxiety attached to all four figures rather than just to the critical figure because of their low discriminability at a low level of awareness. Upon presentation of the stimulus, *S* tensed, gasped, or wrapped a leg around the chair, regardless of which figure was selected. The total conditioned stimulus was probably even more complex, consisting of a stimulus presentation plus a lapse of time of two seconds. Generally, anxiety reduction is assumed to follow upon withdrawal of the cues eliciting the anxiety. However, termination of the exposure period brought no visible end to the *S*'s malaise. It was not until two seconds had elapsed and no shock had occurred that signs of relaxation or a relieved and grateful sigh appeared. In the absence of any good measure of momentary changes in the anxiety level it is reasonable to think anxiety drive reduction and reinforcement of the prior perceptual response occurred at this point.

This interpretation, which holds that the punished response is "weakened" only because competing responses are strengthened is advanced because it has worked well in describing the avoidance learning of skeletal responses (9, 16). Logically, of course, the shift in relative perceptibility of these figures could be credited either to sensitization to the nonpunished stimuli or desensitization to the punished stimuli, or both. The data prove only that at least one of these reactions must have occurred. This ambiguity is shared, however, with studies in which a single stimulus is presented by the ascending method of limits. In studies following the latter design, the assessed strength of the veridical percept is relative to the strength of competing perceptual hypotheses (12). The forced-choice procedure simply delimits the range of competing perceptual responses to a high priority few. Until the evidence for sensitization and desensitization is less equivocal, perceptual vigilance and defense can legitimately be identified only with shifts in balance among competing perceptual responses, in the one case toward a critical percept, and in the other case against it.

This study supports the view that perceptual defense and vigilance are learned reactions to anxiety arousing stimuli. It is consistent with Postman's recent statement that "the facts which give rise to this concept [of perceptual defense] can better be subsumed under other, more general, principles of perception" (12, p. 298); the same could be said for vigilance. Both concepts, it is suggested, may be subsumed under more general principles of perceptual learning, but this statement is not to be construed as asserting that perceptual defense and vigilance are therefore explained away.

Though the mechanisms appear to follow more general principles of learning, they nevertheless retain some identity as mechanisms.

SUMMARY

This study experimentally produces the learning of perceptual defense and vigilance and proposes a behavior theory analysis of the learning process. According to this analysis, perceptual defense is learned when the perceptual response to a threatening stimulus is punished and competing perceptual responses are instrumental to anxiety reduction. Competing perceptual responses when reinforced are strengthened at the expense of the critical perceptual response. Perceptual vigilance is learned when the perceptual response to a threatening stimulus is reinforced by anxiety reduction and competing perceptual responses are punished.

Avoidance learning procedures were employed to change the comparative recognizability of four geometrical figures presented simultaneously below the threshold for awareness. Sixteen *Ss* underwent defense training, another 16 vigilance training. Before each training series, the comparative recognizability of the four figures was assessed for each *S* by a forced-choice technique. In both training series, *S* reported on each trial the position of the figure he found clearest. During defense training, if he selected the critical figure, he was shocked. If he selected one of the other three figures, he avoided the shock. During vigilance training, selection of the critical figure was instrumental to shock avoidance while selection of any of the other three figures incurred the shock. After each training procedure, comparative recognizability of the figures was assessed again. Fourteen of 16 defense *Ss* found the critical figure less recognizable after training than before. Thirteen of 16 vigilance *Ss* found the critical figure more recognizable than before training. Learning in both cases proceeded in the absence of awareness.

REFERENCES

1. BLACKWELL, R. Psychophysical thresholds: experimental studies of methods of measurement. *Eng. res. Bull. No. 36*, Univer. of Michig .n, 1953.
2. BLUM, G. S. An experimental reunion of psychoanalytic theory with perceptual vigilance and defense. *J. abnorm. soc. Psychol.*, 1954, **13,** 94–99.
3. ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61,** 175–182.
4. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41,** 401–410.
5. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58,** 113–123.
6. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19,** 471–482.
7. LAZARUS, R. S. Is there a mechanism of perceptual defense? A reply to Postman, Bronson, and Gropper. *J. abnorm. soc. Psychol.*, 1954, **49,** 396–398.
8. LYSAK, W. The effects of punishment upon syllable recognition thresholds. *J. exp. Psychol.*, 1954, **47,** 343–351.
9. MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem solving." *Harvard educ. Rev.*, 1944, Spring, 102–148.
10. OHRBACH, J. Retinal locus as a factor in recognition. *Amer. J. Psychol.*, 1952, **65,** 555–562.
11. POSTMAN, L. On the problem of perceptual defense. *Psychol. Rev.* 1951, **58,** 113–123.
12. POSTMAN, L. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation: A symposium*. Lincoln: Univer. of Nebraska Press, 1953.
13. POSTMAN, L., BRONSON, W. C., & Gropper, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48,** 215–225.
14. PUSTELL, T. Cue and drive aspects of anxiety in relation to perceptual vigilance and defense. Unpublished doctoral dissertation, Univer. of Michigan, 1955.
15. SIDMAN, M. On the persistence of avoidance behavior. *J. abnorm. soc. Psychol.*, 1955, **50,** 217–221.
16. SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: Acquisition in normal dogs. *Psychol. Monogr.* 1953, **67,** No. 4, (Whole No. 354).
17. SOLOMON, R. L., KAMIN, L. J., & WYNNE, L. C. Traumatic avoidance learning: the outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, **48,** 206–216.

Received February 11, 1957.

THE RELATIVE AVERSIVENESS OF WARNING SIGNAL AND SHOCK IN AN AVOIDANCE SITUATION

MURRAY SIDMAN AND JOHN J. BOREN[1]

Walter Reed Army Institute of Research

IT IS a widespread clinical and personal observation that the anticipation of pain may be worse than the pain itself. Yet the phenomenon has received little experimental attention. Possibly relevant observations are those of Gwinn (2) and Solomon, Kamin, and Wynne (18), who found that extinction of avoidance behavior was not hastened when the avoidance response was punished by shock. Kamin (4) demonstrated that a response which terminates a warning stimulus can be conditioned even when the response is regularly shocked. Escape from the warning stimuli was possibly sufficiently reinforcing to counteract the effects of punishment. Hill, Kornetsky, Flanary, and Wikler (3) found that shock intensities were overestimated by subjects who were not familiarized and reassured with respect to "the potentially fear-inspiring experimental situation" prior to the experiment.

A more direct experimental demonstration would consist of an arrangement that permits objective measurement of the subject's preference, either for a noxious event or for an anticipatory situation prior to the noxious event. The present study demonstrates one such arrangement in the framework of a modified avoidance conditioning procedure. In the classical avoidance experiment, the avoidance response prevents a shock *and* terminates the anticipatory warning signal. Shock avoidance and escape from the anticipatory situation are tied together. In the present modification of the classical technique, the subject can behave in either of two alternative fashions: (*a*) postpone the shock, while at the same time maintaining the warning signal or (*b*) take the shock while simultaneously terminating the warning signal. The shock can be avoided only at the cost of prolonging the conditions that require the anticipatory avoidance behavior. The anticipatory situation can be terminated only at the cost of receiving a shock.

[1] Now at the Merck Institute for Therapeutic Research.

EXPERIMENT I

Subjects and Apparatus

The *S*s were 4 male albino rats. The lever-pressing apparatus, with its associated automatic programming and recording devices, has been described previously (12, 14). In brief, the experimental chamber was a metal box with a grid floor and a sloping plexiglass ceiling. Depression of a lever, which was inserted through one wall of the chamber, constituted the measured avoidance response. A small neon pilot light was employed as the stimulus. The behavior was recorded on magnetic counters and cumulative recorders.

Procedure

Avoidance conditioning. The avoidance response (lever-pressing) was first conditioned by means of a previously described "nondiscriminated" avoidance procedure (11). Brief shocks, of a fixed duration (approximately 0.3 sec.), were delivered through the grid floor every 20 seconds. Each time the animal depressed the lever, however, the shock delivery was postponed an additional 20 seconds, measured from the lever depression. Only the downward movement of the lever postponed the shock; continued holding of the lever did not affect the time of shock delivery.

The interval by which each response delays the shock has been termed the response-shock (RS) interval. A response-shock interval of 20 seconds, with no stimulus light present, is referred to as $(RS20'')^{D}$. The superscript "D" indicates the absence of the stimulus light.

Chained avoidance contingencies under stimulus control. After the rate of avoidance responding became reasonably stable on the shock schedule, $(RS20'')^{D}$, the following procedure was instituted. If the animal permitted 15 seconds to elapse without a lever-pressing response, a stimulus light (L) would come on. Every lever press in the absence of the stimulus, however, postponed the onset of L for an additional 15 seconds. In the absence of L, then, there is a response-light (RL) interval instead of a response-shock (RS) interval. This is referred to as $(RL15'')^{D}$, to indicate that each response in the absence of the light postponed the onset of the light for 15 seconds.

When the animal permitted the light to come on, a different schedule went into effect. If there was no response in the presence of the light a shock occurred within five seconds and the light terminated, reinstating $(RL15'')^{D}$. Each response in the light, however, postponed the shock for five seconds. The light would remain on until the animal permitted the shock to occur. Thus, in the presence of the light there is an RS interval of five seconds, which may be denoted as $(RS5'')^{L}$.

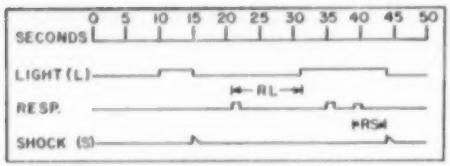

FIG. 1. DIAGRAMMATIC REPRESENTATION OF THE EXPERIMENTAL PROCEDURE (SEE TEXT FOR DETAILS)

The total procedure may, for convenience, be represented symbolically as: $(RL15'')^{D} \rightarrow (RS5'')^{L}, S \rightarrow (RL15'')^{D} \ldots$. This may be read as follows: "In the absence of the light, the response-light interval is 15 seconds. If the animal permits the RL interval to elapse this produces the light, in the presence of which a response shock interval of 5 seconds is in force. Occurrence of a shock terminates the light and reinstates the response-light interval of 15 seconds." Fig. 1 depicts the schedule, $(RL10'')^{D} \rightarrow (RS5'')^{L}, S \rightarrow (RL10'')^{D}$. (In the present experiment, the RL interval was 15 sec.)

Each experimental session lasted six to eight hours, with a given animal running no more frequently than once every alternate day. Responses were recorded separately in the presence and in the absence of the light.

RESULTS

Results for the four *Ss* are presented in Fig. 2. The control curves in the left-hand frames represent the average response rates during the final sessions of avoidance conditioning, in which the prevailing shock schedule was $(RS20'')^{D}$. The curves in the right-hand frames represent the rate of responding in the light and in the absence of the light following the change in procedure. In three of the four cases the response rate in the stimulus light displayed an initial increase over the rate in $(RS20'')^{D}$, but in all cases the rate gradually dropped to a level well below the control rate. The response rate in the absence of the light did not reveal any marked trend from session to session that was consistent among all the *Ss*, but, in all cases, ended up higher than the rate during the warning stimulus.

DISCUSSION

On the basis of previous findings, we might have expected a high rate of avoidance responding in the presence of the warning stimulus and a low rate in its absence. This outcome would have been consonant with the findings of Mowrer and Lamoreaux (8, 9) in the classical avoidance situation and with those of Sidman in a situation in which the avoidance responses could postpone the warning stimulus (14). It has also been shown previously (6, 12, 13) that the rate of avoidance responding is an inverse function of the response-shock interval, within certain limits. On this basis alone, a higher response rate during the warning stimulus would be predicted, when the response-shock interval was only five seconds, than during the control phase, when the response-shock interval was 20 seconds.

In none of the previous experiments, however, was the *S* faced with the dilemma of

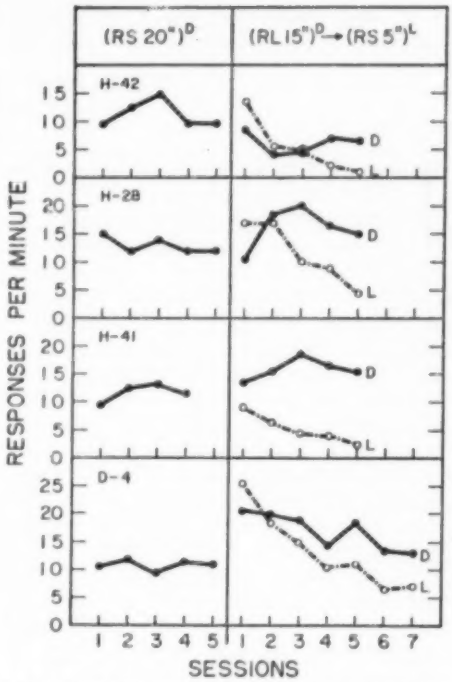

FIG. 2. RATE OF AVOIDANCE RESPONDING IN THE LIGHT (L) AND IN THE ABSENCE OF THE LIGHT (D) FOR 4 ANIMALS

The curves on the left are for the nondiscriminated avoidance procedure, and the curves on the right are for the chaining procedure. Session 1 on the chaining procedure followed immediately after the last session on the nondiscriminated procedure.

choosing between the warning stimulus and the shock. The response that avoided the shock was also responsible for termination of the warning signal. The present case, in which such a choice was required, produced behavior whose pattern was diametrically opposite to that which is customarily observed. The rate of avoidance responding was low in the presence of the warning stimulus and high in its absence. This was the pattern to be expected if the shock was preferable to the signal. For while lever pressing during the warning stimulus did succeed in postponing the shock, it also served to prolong the stimulus. Behavior other than lever pressing, however, was the only means of terminating the warning stimulus. The low rate of lever pressing indicates that other behavior did, indeed, become prepotent, even though an inevitable shock was one of the consequences of such behavior.

EXPERIMENT II

If the decline in the rate of lever-pressing during the stimulus was due to the reinforcement of competing behavior through the termination of the stimulus, it should be possible to restore the lever-pressing rate by terminating the light independently of any behavior. This was accomplished by permitting the stimulus light to remain on for five minutes, regardless of whether or not the animal pressed the lever.

Subjects and Apparatus

The *Ss* were five male albino rats. The apparatus was the same as in Experiment 1, except for the shock power supply. In this experiment an Applegate shock generator set at maximum output was used. At this setting, the shock varied from 3 to 5 ma.

Procedure

Shock terminates stimulus. This procedure was the same as the "chained avoidance contingencies under stimulus control" procedure described in Experiment I. The salient feature here is the termination of the stimulus coincidentally with the delivery of the shock. All animals were run on the schedule, $(RL20'')^D \rightarrow (RS10'')^L$, $S \rightarrow (RL20'')^D$ . . . for a total of 100 hours. Rat AA-1 was also run on the schedule, $(RL20'')^D \rightarrow (RS20'')^L$, $S \rightarrow (RL20'')^D$ . . . for 100 hours. Rat GF-81 was also run for only one session on the schedule, $(RL20'')^D \rightarrow (RS5'')^L$, $S \rightarrow (RL20'')^D$. . . . Unfortunately, this animal died on the day following this session, but his data were included with the others.

Five-minute stimulus duration. On this procedure, whenever the animal permitted the stimulus to appear it remained on for five minutes, regardless of the number of responses emitted or the number of shocks delivered. Termination of the stimulus was based upon the passage of time alone, and was independent of the animal's behavior. In terms of our symbolic notation, this procedure might be expressed as $(RL20'')^D \rightarrow (RS10'')^L$, $(5\text{-min.})^L \rightarrow (RL20'')^D$. . . .

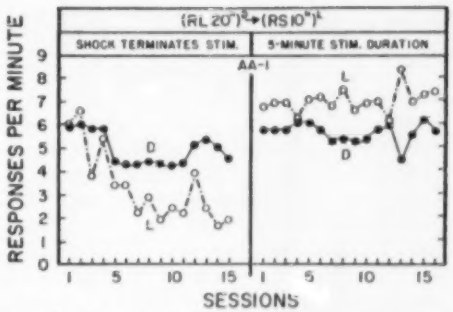

FIG. 3. RATE OF AVOIDANCE RESPONDING IN THE LIGHT (L) AND IN THE ABSENCE OF THE LIGHT (D) FOR RAT AA-1. THE SESSIONS ARE CONSECUTIVE

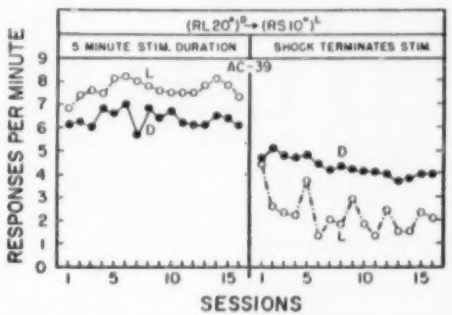

FIG. 4. RATE OF AVOIDANCE RESPONDING IN THE LIGHT (L) AND IN THE ABSENCE OF THE LIGHT (D) FOR RAT AC-39. THE SESSIONS ARE CONSECUTIVE

The RL and RS intervals for each animal were the same as those described above in the "shock terminates stimulus" procedure. Every animal was run for 100 hours on each schedule. On the procedure involving $(RL20'')^D \rightarrow (RS10'')^L$, rats GF-81 and AA-1 ran first on the "shock terminates stimulus" procedure, after which they were changed to the "five-minute stimulus duration" procedure. All the other animals, along with GF-81 and AA-1 on the procedures that involved $(RS5'')^L$ and $(RS20'')^L$, respectively, ran on the "five-minute stimulus duration" procedure first.

RESULTS

Fig. 3 and Fig. 4 give a session-by-session picture of the response rates under both pro-

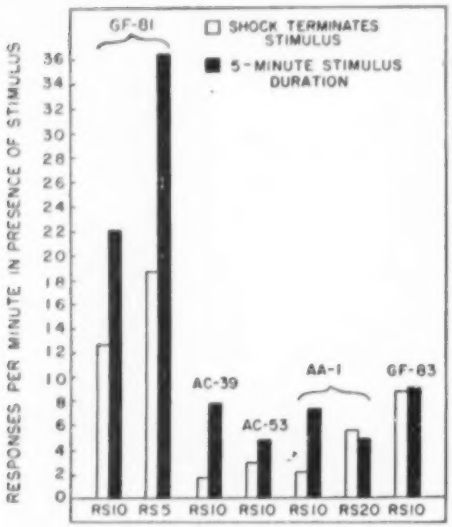

FIG. 5. RATE OF AVOIDANCE RESPONDING IN THE PRESENCE OF THE LIGHT WHEN THE LIGHT TERMINATES WITH THE SHOCK AND WHEN THE DURATION OF THE LIGHT IS INDEPENDENT OF THE SHOCK

The RL interval is 20 sec. in all cases.

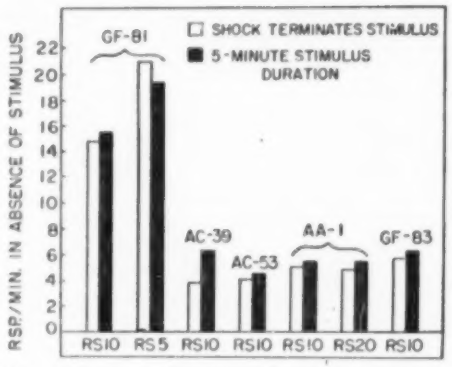

FIG. 6. RATE OF AVOIDANCE RESPONDING IN THE ABSENCE OF THE LIGHT WHEN THE LIGHT TERMINATES WITH THE SHOCK AND WHEN THE DURATION OF THE THE LIGHT IS INDEPENDENT OF THE SHOCK

The RL interval is 20 sec. in all cases.

cedures for two of the animals. Rat AA-1, which started on the "shock terminates stimulus" procedure, replicated the data presented in Experiment I. The response rate in the presence of the stimulus fell off relative to its own initial level and to the rate in the absence of the stimulus. When the stimulus duration was fixed at five minutes and was, therefore, independent of the animal's behavior, there was a marked increase in the rate of responding during the stimulus. Rat AC-39, started on the "five-minute stimulus duration" procedure, showed essentially the same results. When the stimulus termination was made dependent upon failure to press the lever and receipt of a shock, the response rate in the stimulus dropped markedly. In both animals there was only a slight change in response rate in the absence of the stimulus.

A summary of the data for all animals is presented in Fig. 5. This figure shows the median response rate in the stimulus over the last four sessions of each procedure for each animal. In all cases the RL interval was 20 seconds. The RS intervals employed are indicated beneath the corresponding bars of the graph. Of the five cases in which the RS interval was 10 seconds, the rate during the stimulus was higher when the stimulus duration was independent of the animal's behavior. In the case of GF-83, however, the difference was too small to be taken seriously, and this animal must be considered an exception.

When rat GF-81 was run on an RS interval of five seconds there was again a striking difference in rate between the two procedures. It must be remembered, however, that GF-81 had only one session on $(RS5'')^{L}$ and that the rates depicted here are undoubtedly not the final levels that would have been reached if additional sessions could have been run. A most interesting finding is that yielded by Rat AA-1 when the RL and RS intervals were both 20 seconds. In this case, it made little difference whether or not the stimulus duration depended upon the *S*'s behavior. Response rates in the presence of the stimulus were essentially the same in both procedures.

The effect of the procedural differences upon the response rate in the absence of the stimulus is shown in Fig. 6. With one exception, the rate on the five-minute stimulus duration procedure was higher, but in no case were the differences for a given animal as great as those displayed by the response rates in the presence of the stimulus depicted in Fig. 5. The one exception, GF-81 at an RS interval of five seconds, probably does not represent a true stable level

of responding because only one session was included.

DISCUSSION

These data substantiate the view that the decline in lever-pressing rates during the stimulus, observed in Experiments I and II, is due to the reinforcement of nonlever-pressing behavior by the termination of the stimulus; this, in spite of the fact that a shock accompanied stimulus termination. When the stimulus duration was fixed at five minutes, regardless of the animal's behavior, the response rate in the presence of the stimulus increased. In this procedure it was impossible for the animal to "take the shock in order to terminate the stimulus," and the lever-pressing rate assumed a value consistent with the prevailing response-shock interval.

Most current explanations of discriminated avoidance behavior attribute the maintenance of the avoidance response to a "drive" or "aversive" function acquired by the warning stimulus as a result of its temporal relation to the shock (e.g., 5, 7, 10, 17). Termination of the stimulus is held to provide reinforcement for the avoidance response. The present data seem to indicate that the stimulus-shock pairings actually make the warning stimulus more aversive than the shock from which it derived its function. But the data of Rat AA-1 do not support the suggestion that the aversive function of the warning stimulus derives simply from its association with the shock. Rather than dismiss the performance of AA-1 as an example of unexplained variability, we may view the apparent exception with greater than ordinary interest. Rat AA-1 displayed the expected difference in responding under the two procedures when the RL interval was 20 seconds and the RS interval was 10 seconds. There was no large difference, however, when both intervals were equal. In spite of the fact that the warning stimulus was inevitably paired with shock, there is little indication that the stimulus was more aversive than the shock when RL and RS intervals were each 20 seconds. Some factor in addition to the stimulus-shock pairings must be involved in determining the relative aversiveness of stimulus and shock.

Sidman and Boren have suggested (16) that the discriminated avoidance situation may be considered as a "multiple schedule" (1) in which one set of avoidance contingencies prevails in the presence of the warning stimulus and another in its absence. In the present experiments, the two sets of avoidance contingencies may be specified in terms of the RL and RS intervals, the latter prevailing in the presence of the warning stimulus, and the former in its absence. When both intervals were 20 seconds, there was no great advantage to the animal in taking the shock and terminating the stimulus, for the temporal contingencies were the same in the presence and absence of the stimulus. A relatively high rate of lever pressing was therefore maintained in the stimulus, and the shift to a fixed stimulus duration produced no appreciable change. (The fact that this situation obtained when both intervals were numerically equal may be regarded as fortuitous. It is possible that for other animals a much larger difference between the RL and RS intervals would be required to produce a decline in rate during the stimulus.) Other experiments have served to confirm this explanation. It was found that the aversiveness of the warning signal is a function of the RS interval that prevails in the presence of the signal, of the RL interval that prevails in its absence, and of the difference between the two (15). It is necessary to take into account not only the anticipatory situation from which the *S* escapes, but also the conditions that obtain subsequent to the escape. If the latter are relatively favorable, the anticipatory signal may actually become more aversive than the shock itself.

SUMMARY

An avoidance situation was arranged in which the animals could either (*a*) postpone the shock and prolong the warning signal or (*b*) take the shock and terminate the warning signal. With this procedure, we observed a low rate of avoidance responding in the presence of the warning signal and a high rate in its absence. The animals behaved in such a way as to terminate the stimulus as quickly as possible, even though a shock accompanied each stimulus termination. When the signal duration was made independent of the animal's behavior, the rate of avoidance responding in the stimulus increased. It was suggested that the avoidance contingencies in the presence and absence of the warning stimulus are important determiners of the relative aversiveness of stimulus and shock.

REFERENCES

1. FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* Appleton-Century-Crofts, in press.
2. GWINN, G. T. The effects of punishment on acts motivated by fear. *J. exp. Psychol.*, 1949, **39,** 260–269.
3. HILL, H. E., KORNETSKY, C. H., FLANARY, H. G., & WIKLER, A. Effects of anxiety and morphine on discrimination of intensities of painful stimuli. *J. clin. Investig.*, 1952, **31,** 473–480.
4. KAMIN, L. J. The effects of termination of the CS and avoidance of the US on avoidance learning. *J. comp. physiol. Psychol.*, 1956, **49,** 420–424.
5. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
6. MORSE, W. H., & HERRNSTEIN, R. J. The maintenance of avoidance behavior using the removal of a conditioned positive reinforcer as the aversive stimulus. *Amer. Psychologist,* 1956, **11,** 430. (Abstract)
7. MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem-solving." *Harv. educ. Rev.*, 1947, **17,** 102–148.
8. MOWRER, O. H. & LAMOREAUX, R. R. Avoidance conditioning and signal duration: a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, **54,** No. 5 (Whole No. 247).
9. MOWRER, O. H., & LAMOREAUX, R. R. Conditioning and conditionality (discrimination). *Psychol. Rev.*, 1951, **58,** 196–212.
10. SCHOENFIELD, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. H. Hoch & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950. Pp. 70–99.
11. SIDMAN, M. Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science,* 1953, **118,** 157–158.
12. SIDMAN, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *J. comp. physiol. Psychol.*, 1953, **46,** 253–261.
13. SIDMAN, M. On the persistence of avoidance behavior. *J. abnorm. soc. Psychol.*, 1955, **60,** 217–220.
14. SIDMAN, M. Some properties of the warning stimulus in avoidance behavior. *J. comp. physiol. Psychol.*, 1955, **48,** 444–450.
15. SIDMAN, M. Conditioned reinforcing and aversive stimuli in an avoidance situation. *Trans. N. Y. Acad. Sci.*, 1957, **19,** 534–544.
16. SIDMAN, M., & BOREN, J. J. A comparison of two types of warning stimulus in an avoidance situation. *J. comp. physiol. Psychol.*, in press.
17. SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.
18. SOLOMON, R. L., KAMIN, L. J., & WYNNE, L. C. Traumatic avoidance learning: the outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, **48,** 291–302.

Received February 14, 1957.

SELECTIVE RECALL IN SCHIZOPHRENIA AND ITS RELATION TO EGO STRENGTH[1]

ARTHUR S. TAMKIN[2]

Duke University

THE PURPOSE of this study is to investigate the nature of selective recall in schizophrenic patients. The hypotheses offered are that schizophrenics manifest a distinctive type of memory process in contrast to normal subjects, resulting in the preferential remembering of experiences connoting personal failure and diminished self-esteem, and that this process may be related to deficient ego strength.

This characteristic memory process may be inferred both from clinical observations of schizophrenics and from theoretical considerations of schizophrenia. Certain aspects of its symptomatology, suggesting selective memory functioning, have been described as brooding preoccupations (4), chronic feelings of insecurity and inadequacy (20), and reaction sensitivity to events which reduce self-esteem (6). Cameron has explained this symptomatology as follows: "In the childhood of both paranoid and schizophrenic patients . . . there have been developed self attitudes which leave the individual unusually reaction sensitive to signs that he may be guilty, inferior, or unworthy in the eyes of others" (5, p. 489).

Although in clinical descriptions and theoretical discussions pathological selective remembering has been designated by different terms, the common implication is that the direction of thinking and remembering in the schizophrenic is determined by his peculiar feelings, needs, and emotional complexes. It seems that he does not passively absorb impinging stimuli and recall his experiences indiscriminately; rather, directive intrapsychic forces predispose him to respond to and retain material which is congruent with his pathological motivations.

In normal functioning, selectivity in both perception and memory has been demonstrated by a variety of experimental methods, and it has been related to basic needs and defensive mechanisms of the personality. Personal needs and values, for example, have been shown to enhance the perceptual recognition of need- and value-related words (14), and the preferential perception of tasks associated with failure and success has been shown to be determined by individual defensive mechanisms (9, 15). Studies of selectivity in memory are represented by a large body of literature reviewed in several publications (3, 16). The general findings which have emerged are that memory normally functions in a selective manner, enhancing retention of material with which the person is in agreement (7), which is congruent with his value-orientation (8), and which has pleasant affective connotations (12).

A method for measuring selective recall which has enjoyed wide use is the recall of completed and incompleted tasks, introduced by Zeigarnik (22) in 1927. This method is quite suitable for the purposes of this investigation, for when it is applied as a simulated intelligence test in order to enhance the ego involvement of the subjects, the experimenter may systematically manipulate the subjects' experience of self-esteem through success and failure on the tasks. Moreover, motives related to the defense of the ego or other psychologically significant adaptive responses may be evoked. The method of completed and incompleted tasks creates a realistic situation which maintains the subjects' interest by simulating a familiar kind of competitive situation, and is, therefore, a convincing experimental analogue of an emotionally arousing experience which has real personal significance for the participants. Alper (2) has demonstrated that when the completed and incompleted task method is used under ego-involving conditions, individuals who recall a preponderance of successes in this threatening situation are characterized by a "strong ego," whereas those who recall a preponderance of failures have a "weak ego." Several subsequent studies have shown that

[1] This paper is based on a Ph.D. dissertation at Duke University. The author wishes to thank Drs. Louis D. Cohen, George H. Collier, and Gregory Kimble for their invaluable assistance on this project, and Dr. Burke M. Smith of the Roanoke Veterans Administration Hospital who made available patients and facilities.

[2] Now at the Veterans Administration Hospital, Northampton, Mass.

normally the completed tasks are selectively retained (1, 11, 17), and others have confirmed the relationship between this tendency and level of ego strength (9, 10, 13).

When completed and incompleted tasks are presented to schizophrenics, one may hypothesize (*a*) preferential recall of the incompleted puzzles and (*b*) a relation between this tendency and deficient ego strength. These specific hypotheses were tested by the experiment reported here.

METHOD

In order to test the hypothesis that schizophrenics manifest selective recall of failure experiences, their recall of completed and incompleted puzzles which were administered ostensibly as an intelligence test was analyzed. An incidental rather than an intentional (set-to-learn) learning situation was utilized because of its superiority for the demonstration of selective remembering (1). The condition of ego involvement, created by simulated intelligence testing, was established in order to enhance the probability that incompleted puzzles would connote failure, thus presenting a threat to self-esteem. The second hypothesis, which states that the selective recall of schizophrenics is related to deficient ego strength, was explored with the Scott-Duke Questionnaire (18), designed to measure the trait of self-acceptance, a personality characteristic presumably related to level of ego strength. An inverse relation between scores on the Self-acceptance scale of this questionnaire and a preponderant recall of incompleted puzzles would be considered confirmative of the second hypothesis.

Subjects

An experimental and a control group of 24 *Ss* each were used in the experiment. The experimental group was composed of male schizophrenics, patients at the Veterans Administration Hospital, Roanoke, Virginia, randomly selected on the basis of availability for psychological testing and ability to cooperate on the testing procedures. Four *Ss* had to be eliminated because of failure to cooperate. The diagnoses were made by the psychiatric staff. All patients used were recent admissions who had received no psychiatric treatment during their current hospitalization at the time of the experiment. The control group consisted of 24 adult males who had never received psychiatric treatment and who claimed that they did not require such treatment. They were considered to represent a group of normal individuals. They were selected according to criteria of color, age, education, sex, and geographical residence so as to yield a group comparable to the experimental group for these personal variables. The data are presented in Table 1.

TABLE 1
VARIABLES ON WHICH THE SCHIZOPHRENIC AND CONTROL GROUPS WERE MATCHED

Variable	Schizophrenics		Controls		*t* test
	N	Mean	*N*	Mean	
Education	—	8.92	—	9.25	.43[a]
Age	—	30.08	—	29.75	.18[a]
Negro	4		4		
White	20		20		

Note.—Subjects of both groups resided in the same locality and all were males.

[a] p is greater than .20.

The members of the control group were obtained from three separate sources in response to a solicitation to donate one hour of time to an unspecified research project. Eleven *Ss* volunteered at a labor union meeting, ten volunteered at a state employment office for a monetary incentive, and three were obtained while waiting for relatives at an outpatient clinic.

Materials

The materials for the completed and incompleted tasks consisted of 10 jigsaw puzzles, each depicting a common animal with its name printed conspicuously along the border. The animals were kangaroo, giraffe, parrot, bear, rooster, eagle, swan, tiger, geese, and elephants.

The Scott-Duke Questionnaire comprises the *L* and *K* scales of the MMPI, which are used as a check on validity, and a Self-acceptance scale. It is a self-administered, true-false type of questionnaire, scored by means of an objective key. According to Scott (18), the Self-acceptance scale was based upon Sheerer's (19) definition of the self-accepting person, and it was shown to have a high reliability and validity as a measure of self-acceptance. The split-half reliability was reported by Scott as .94. Its validity was predicated upon its correlation of .77 (.83, corrected for attenuation) with Rogers' Q-sort technique with 58 *Ss*. The empirical finding that it differentiated between a variety of psychiatric disorders ranging from the psychoneuroses to schizophrenia, which yields the lowest score, as well as its relation to the Q-sort measure of self-acceptance, tends to justify its use in this experiment as an estimate of ego strength.

Experimental Procedure

All *Ss* in both the experimental and control groups underwent an identical procedure, but only 17 *Ss* from each group, matched for the personal variables, received the Scott-Duke Questionnaire. The procedure consisted of the following sequence of individually administered tests: a vocabulary test, assembly of jigsaw puzzles, digit span test, test for recall of jigsaw puzzles, and the Scott-Duke Questionnaire. The vocabulary and digit span tests were introduced in order to simulate more plausibly an intelligence testing condition, and

the instructions given to both groups were worded appropriately. The ten jigsaw puzzles, constituting the memory material, were administered in a random sequence. The order of completions (C) and incompletions (I) was identical for each *S* in the sequence ICCCIICIIC. The *S*s, timed for their performance on each puzzle, were allowed to complete every one designated as a C puzzle, and were prevented from completing every I puzzle by the imposing of an arbitrary time limit just before the last two pieces were fitted into the partially assembled puzzle. After completing each C puzzle, the *S* was asked to tell what the picture was if he did not do so spontaneously. In order to ensure equal exposure time to all puzzles, the I puzzles, which were finally completed by the experimenter, were exposed to the *S*'s view, and he was asked to state what the picture was. Immediately after the interpolated digit span test, each *S* was asked to recall the names of all the jig-saw puzzles which he could remember.

TABLE 2
FREQUENCIES OF SUBJECTS DEMONSTRATING EACH TYPE OF RECALL PATTERN

	I	Equal	C
Schizophrenics	15	4	5
Controls	6	9	9

TABLE 3
SUMMARY OF ANALYSIS OF VARIANCE OF SELF-ACCEPTANCE SCORES

Source of Variance	*df*	Sum of Squares	Mean sum of Squares	*F*	*P*
Between groups	2	31.68	15.84	.32	> .05
Within groups	31	1544.08	49.81		
Total	33	1575.76			

RESULTS

The chi-square test was used to answer the question whether any distinctive pattern of recall was associated with the schizophrenics as compared with the control *S*s. The *S*s of each group were divided into three categories according to their recall patterns: (*a*) the recall of a preponderance of I puzzles (I), (*b*) recall of a preponderance of C puzzles (C), and (*c*) recall of an equal number of I and C puzzles (Equal). The first two patterns represent different types of selective recall, while the third pattern shows no selectivity. The distribution of *S*s falling into each category is shown in Table 2. Chi square for the association of recall pattern with psychiatric status (normal or schizophrenic) was 6.89 ($.05 > p > .02$). It may be observed that while a majority of the schizophrenic group demonstrated the I pattern, a minority of the control group also did so. The majority of the control group was distributed evenly between the Equal and C patterns.

In order to determine the relationship between scores on the Self-acceptance scale and the three recall patterns, all *S*s were sorted according to their recall patterns and an analysis of variance of their Self-acceptance scores performed. The obtained *F* of .32, as shown in Table 3, is not significant.

Since Self-acceptance scores were unrelated to the recall patterns, evaluations were made of the reliability and applicability of the scale in the assessment of ego strength with the *S*s of this experiment. Internal consistency was estimated by the split-half method, using the Spearman-Brown formula. A coefficient of .90 was obtained for the 34 *S*s in the combined groups. This coefficient, similar to that of .94 reported by Scott, is high enough to suggest that the scale has high internal consistency. While a more meaningful evaluation of reliability would involve the question of the stability of the measurements over time, test-retest data were not available to permit this type of evaluation. With regard to the applicability of the Self-acceptance scale in estimating ego strength, on a priori grounds it should differentiate between hospitalized schizophrenics and normal persons. The mean Self-acceptance score of the control group was 30.65, and of the schizophrenic group 25.12. A *t* of 2.96 was significant at the .01 level, but since there was considerable overlap between the two distributions, its effectiveness in discrimination was rather low.

DISCUSSION

While the hypothesis that schizophrenics selectively recall experiences connoting personal failure and diminished self-esteem was supported by the experimental findings, the hypothesis of the relation of this memory process to deficient ego strength was unsupported. The failure to confirm the second hypothesis may be attributable, in part, to two possible sources of error: one having to do with deficiencies in the instrumentation employed, the other involving certain weaknesses of the concept of ego strength.

The estimation of ego strength by means of

the Self-acceptance scale seemed tenable in view of its demonstrated relationships to psychopathology and to an independent, objective measure of self-acceptance. However, its power of discrimination proved to be quite low owing to the degree of overlap of scores in the schizophrenic and normal groups. Furthermore, while high split-half reliability was demonstrated, the question of the stability of measurement over time was not evaluated owing to the unavailability of test-retest data. This would seem to be the more important consideration in establishing the level of reliability because of the probable susceptibility of the schizophrenics to transient influences and momentary moods which would make their responses to the Self-acceptance scale undependable. In view of these sources of experimental error, one may question the effectiveness of the Self-acceptance scale as an estimate of ego strength.

Another factor which might be responsible for the finding of no relationship, in spite of the positive findings of other experimenters who used different instruments and normal *S*s, is that due to ambiguous operational and conceptual definitions of ego strength, different facets of this trait are measured by the various instruments employed. The findings of a recent study comparing several independent measures of ego strength have drawn attention to this ambiguity on empirical grounds (21). Because ego strength, defined by Eriksen as "the individual's capacity for appraising the reasonable limits in his interpretations and perceptions of his environment" (10, p. 46) and measured by Rorschach $F+\%$, has been shown to correlate significantly with selective recall, this is no indication that ego strength measured by the Self-acceptance scale should. This consideration argues the need for a more precise definition of the presently vague concept of "ego strength." Certainly, the negative finding obtained in this experiment points up the need for continued research on the problems of ego strength and its role in schizophrenia.

SUMMARY

Two hypotheses concerning certain aspects of schizophrenic symptomatology were tested: (*a*) Schizophrenics manifest a distinctive type of memory process which results in the selective recall of experiences connoting personal failure and diminished self-esteem and (*b*) this process is related to deficient ego strength. The method used to test the first hypothesis was to administer jigsaw puzzles as a simulated intelligence test, allowing each *S* to complete only half of the group of puzzles, and to determine whether there occurred selective recall by the schizophrenics for the incompleted puzzles connoting personal failure. As a test for the second hypothesis the Scott-Duke Questionnaire, which contained a Self-acceptance scale and two validity scales from the MMPI, was administered to provide a measure of ego strength, and its power of discrimination between *S*s grouped according to their recall pattern was ascertained. The *S*s consisted of 24 schizophrenic patients who were recently admitted or readmitted for hospitalization and a control group of 24 normal adults matched for age, education, sex, color, and geographical residence. While the first hypothesis was confirmed at a significant level of confidence, the second was not supported by the experimental data. The failure to confirm the second hypothesis seemed to be attributable, in part, to two sources of error; one arising from deficiencies in the instrumentation, and the other from certain weaknesses of the concept of ego strength. It is conceivable that due to the vague operational and conceptual definitions of ego strength, different facets of this trait may be measured by the various instruments which have been employed. It is believed that the negative finding concerning the second hypothesis points up the need for continued research on the problem of ego strength and its role in schizophrenia.

REFERENCES

1. ABORN, M. The influence of experimentally induced failure on the retention of material acquired through set and incidental learning. *J. exp. Psychol.*, 1953, **45**, 225–231.
2. ALPER, THELMA G. Memory for completed and incompleted tasks as a function of personality: an analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403–420.
3. BARRETT, DOROTHY M. Memory in relation to hedonic tone. *Arch. Psychol.*, 1938, No. 223.
4. BLEULER, E. *Dementia praecox or the group of schizophrenias.* (Translated by Joseph Zinkin.) New York: International Universities Press, 1950.
5. CAMERON, N. *The psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.
6. CAMERON, N., & MAGARET, ANN. *Behavior pathology.* Boston: Houghton Mifflin, 1951.

7. CLARK, K. B. Some factors influencing the remembering of prose material. *Arch. Psychol.*, 1940, No. 253.
8. COFER, C., & HAVRON, D. Influence of personal values on learning. *Amer. Psychol.*, 1950, **5**, 274. (Abstract)
9. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230–235.
10. ERIKSEN, C. W. Psychological defenses and "ego strength" in the recall of complete and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954., **49**, 45–50.
11. KENDLER, TRACY S. The effect of success and failure on the recall of tasks. *J. gen. Psychol.*, 1949, **41**, 79–87.
12. LAFFAL, J. The learning and retention of words with association disturbances. *J. abnorm. soc. Psychol.*, 1952, **47**, 454–462.
13. LAZARUS, R. S., & LONGO, N. The consistency of psychological defenses against threat. *J. abnorm. soc. Psychol.*, 1953, **48**, 495–499.
14. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.
15. POSTMAN, L., & SOLOMON, R. L. Perceptual sensitivity to completed and incompleted tasks. *J. Pers.*, 1950, **18**, 347–357.
16. RAPAPORT, D. *Emotions and memory*. New York: International Universities Press, 1950.
17. RUSSELL, W. A. Retention of verbal material as a function of motivating instructions and experimentally induced failure. *J. exp. Psychol.*, 1952, **43**, 207–216.
18. SCOTT, H. A. The empirical assessment of self-acceptance. Unpublished doctoral dissertation, Duke Univer., 1953.
19. SHEERER, ELIZABETH. An analysis of the relation between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169–175.
20. SULLIVAN, H. S. Research in schizophrenia. *Amer. J. Psychiat.*, 1929, **9**, 553–567.
21. TAMKIN, A. S. An evaluation of the construct validity of Barron's ego-strength scale. *J. clin. Psychol.*, 1957, **13**, 156–158.
22. ZEIGARNIK, B. Das Behalten erledigter und unerledigter Handlungen. *Psychol. Forsch.*, 1927, **9**, 1–85.

Received February 4, 1957.

EFFECTS OF PERCEIVED DEPENDENCY RELATIONSHIPS UPON CONFORMITY TO GROUP EXPECTATIONS[1]

LEONARD BERKOWITZ

University of Wisconsin

TWO PREVIOUS papers have shown that (*a*) groups evaluated as a unit become more highly motivated than groups whose members are evaluated separately (1), and (*b*) the individually evaluated groups that develop the highest task motivation are those whose members initially are highly motivated and then evaluated unfavorably (2). In both cases it was suggested that the final high task motivation resulted from a shared perception of group member interdependence for reward attainment. The present experiment provides a direct test of this hypothesis.[2] In addition, an attempt will be made to demonstrate that the hypothesized relationship is a special case of a more general relationship between perceived member interdependence and motivation to conform to relevant group expectations.

Several components of perceived interdependence must be considered in accounting for its effect upon group-task motivation. The writer has proposed elsewhere (2) that the degree to which an individual feels dependent upon others is a joint function of two factors: the judged probability that his progress toward his goal is dependent upon the activity of these others (perceived instrumentality) and the extent to which he values this goal.[3] In all likelihood, perceived instrumentality and goal value affect the degree of perceived dependence regardless of the particular direction of the dependency relationship. Thus, assuming that each group member's task performance is necessary for any single member's goal attainment, a subject (*S*) should see himself as being highly dependent upon his teammates when he is likely to gain a goal he values highly through group activity with them. He also should perceive his teammates as being highly dependent upon him when the group activity gives them access to a goal they would value highly. Any given member of an *interdependent* group probably has both of these perceptions; namely, that he is dependent upon the others for the attainment of his goals, and also that they are dependent upon him for their goal attainment.

The primary motive for an individual who can attain his task-associated goal independently of the others in the group (i.e., an *S* who perceives low instrumental interrelationships among the group members) stems solely from the value of this goal to him. The member of an instrumentally interrelated group, however, has additional motives created by the

[1] This study was supported by funds granted by the University Research Committee of the University of Wisconsin. The author would like to acknowledge Mr. Sidney Willmuth's invaluable assistance in the collection of the data.

[2] There is some evidence providing tentative support for the hypothesis. Deutsch (3) has shown, for example, that cooperative groups are more highly task motivated than competitive groups. The individuals in the cooperative groups are described as being "promotively interdependent" in that any member's instrumental behavior could advance all members toward the goal. Deutsch's results would appear to indicate, then, that these interdependent *S*s were more highly motivated toward the assigned group task than the less promotively interdependent members of the competitive groups. However, the obtained differences may not be due to the degree of perceived promotive interdependence per se. There is a possibility that the high group-task motivation was produced by one or more of the by-products of the cooperative relationship, such as the high group cohesiveness, either directly or in interaction with the perceptions of member interdependence.

A somewhat similar difficulty is presented by a second related study. Pepitone (8) found that *S*s who had been told their jobs were important to the group were more highly task motivated than *S*s who had been told their jobs were of relatively little importance. Since group members generally are dependent upon the behavior of those with important jobs, the *S*s' task motivation apparently was directly affected by their belief that others in the group were dependent upon them. However, there also are alternative possibilities. It is conceivable that ego-enhancing motives were aroused by *E*'s description of *S*'s job as being important to the group, and that these motives affected Pepitone's results.

[3] Helen Peak has made a similar analysis of attitudes (7). She has suggested that the position any person takes on an attitude scale is some function of: (*a*) "The judged probability that the [attitude] object leads to good or bad consequences [i.e., the perceived instrumental relationship between the object and the perceiver's goal], and (*b*) the intensity of the effect expected from those consequences." P. 154.

knowledge that others are dependent upon his task behavior for their rewards.

This knowledge may affect the person's task motivation in at least two different ways. First, if the others are in a position to become aware of his performance, he may feel that effective task behavior on his part is necessary to win their good will. Although the desire to win the approval of others probably is an effective motivating force in itself, this desire should be particularly important when the person is dependent upon the others' task activity for his access to a valued goal, as is the case in interdependent groups. Second, his effective performance may be necessary in order to conform to his interiorized standards of proper social conduct. Since the group members in the experiment to be described observed each other's behavior, it is not possible to test these alternatives unequivocally. However, whatever the reason, under the present conditions the task motivation of *S*s should be affected by the perception that others in their group are dependent upon them.

Several specific hypotheses can be formulated on the basis of the foregoing discussion. The first has to do with the perceived instrumental interrelationships among the group members. 1. *Members of groups who are instrumentally interrelated for reward attainment become more highly motivated toward the group task than* Ss *who can achieve their goals independently of the others in the group.* The remaining hypotheses are based upon the assumption that the value of the goal to *S* and/or his teammates also affects the perceived dependency relationships. In highly interdependent groups the task-associated goal is likely to be valued highly by all group members. 2. *Under conditions of instrumental interdependence,* Ss *who are told that both they and their teammates can attain a valued goal through the group activity become more highly task motivated than* Ss *who are told that either only* S *or only* S*'s partner can gain this goal.* At least part of *S*'s task motivation in an interdependent group stems from his knowledge that others are dependent upon him for their attainment of a goal they would value highly. 3. *Given that the instrumentally interrelated group members are aware of each other's behavior,* Ss *instructed that only their partners have access to a valued goal through the group activity become more highly task motivated than* Ss *working for a goal that is less valued by both themselves and their partners.*

METHOD

Group Task and Measure of Task Motivation

Two *S*s, both male or female, were scheduled together, ostensibly to work on a task assessing their "effective intelligence." They were told that the object of the task was for each of them to transmit a list of nonsense syllables to the other as quickly as possible by means of a simulated telegraph system. Each *S* operated two telegraph keys controlling light signals on his partner's desk, "dot" for one light, and "dash" for the other. Employing a telegraphic code supplied to each *S*, *S-I* sent the first nonsense syllable on his list to his partner, *S-II*, who decoded and copied the message. *S-II* then sent the first syllable on his list to *S-I*, who in turn decoded and wrote out the partner's message before sending his second syllable to *S-II*. This alternation was continued for 15 syllables by each *S*. The group members obviously were instrumentally interrelated because of the structure of the task.

After completing a given nonsense syllable, each *S* flashed a third light on his partner's desk. A trial is defined as the time in seconds taken by a given *S* to receive and decode his partner's message and then send his own message. The 15 trials were divided into three blocks of five trials each for computational convenience. The *E* told the *S*s that the first five trials were "for practice" and would not count toward their score. The total time for these trials in the first block provided for each *S* a baseline activity rate from which the time scores for the two remaining blocks were subtracted to yield two improvement scores. The latter scores, the dependent variable in this study, serve as an index of *S*'s motivation to work rapidly in the given block, i.e., his motivation to achieve the designated task objective. Analyses of the data disclose no significant relationship between an *S*'s improvement scores and those of his partner.

Research Design and Experimental Manipulations

Under the pretext that he was being interviewed for life history data, each *S* was individually given the task instructions appropriate to the condition to which he was assigned. Five conditions were established as follows:

(*a*) Both *S* and his partner eligible for a prize, high instrumental interrelationship.
(*b*) *S* alone eligible for a prize, high instrumental interrelationship.
(*c*) *S*'s partner alone eligible for a prize, high instrumental interrelationship.
(*d*) No prizes to be attained, high instrumental interrelationship.
(*e*) Both eligible, low instrumental interrelationship.

In each group, *S-I* (randomly chosen) was informed that ten-dollar prizes were to be awarded to the best *S*s. Accuracy of transmission was listed as one of the criteria, but speed was emphasized. For Condition 1,

S-I was told both he and his partner were eligible for these prizes; the prizes were to be awarded to the best teams. In the second and third conditions, *S-I* was informed that *E* was interested in studying the behavior of only one of the two in the group, and either that he, *S-I*, was the only one being studied and hence eligible (Condition 2) or that his partner was being studied and only he was eligible (Condition 3). *S-II* was never told about the prizes and was instructed only to "try to complete the task quickly." Condition 4 is comprised of these latter *S*s.

The *S*'s eligibility for the prize in these conditions of high instrumental interrelationships is assumed to result in a perceived high degree of dependency upon his partner, while the partner's eligibility presumably causes *S* to view his partner as being highly dependent upon him. The first four conditions therefore can be considered as representing a 2 × 2 factorial design, varying the relations of presumed perceived dependence between *S* and his partner for the prize reward.

In the fifth condition, *S-I* was told that he and his partner would be judged independently and that each was eligible for separate prizes. These instructions were designed to lower the perceived instrumental interrelationships between the group members by counteracting the manifest task structure. A questionnaire, consisting of four seven-point rating scales, was administered at the end of the experimental session to test the success of this manipulation in producing the belief that *S*'s attainment of the reward was not dependent upon the activity of his partner. A sample item was: "How dependent were you upon your partner in completing your task as quickly as possible?" As is indicated in Table 1 below, the items refer to the perceived *instrumental interrelationships* between the team members, and not to the perceived *degree* of dependence resulting from the interaction of this instrumental interrelationship with goal value. Since the four items were significantly and positively intercorrelated, *S*'s perceived instrumentality score was determined by summing his four ratings.

All *S*s were volunteers from liberal arts courses at the University of Wisconsin. There were nine *S*s in each of the conditions. Three *S-I*s were chosen at random from each of the first three conditions and the *S-II*s paired with them were used for the Condition 4 data. Approximately half of the *S*s were females distributed evenly among the conditions.

TABLE 1

PERCEIVED INSTRUMENTAL INTERDEPENDENCY[a]

Condition			
Eligible for Prize	Instrumental Interdependency	Mean	*SD*
1. Both	High	18.89	6.30
2. *S* alone	High	20.56	5.23
3. *S*'s partner alone	High	19.89	4.88
4. No prize	High	20.33	5.66
5. Both	Low	14.00	6.06

[a] $N = 9$ in each condition. There are no significant differences among condition means.

TABLE 2

MEAN TIME IMPROVEMENT SCORES IN CONDITIONS VARYING IN DIRECTION AND DEGREE OF DEPENDENCY BECAUSE OF DIFFERENCES IN PRIZE ATTAINABILITY

	Prize Attainable By *S*		Prize Not Attainable By *S*	
Trials	Prize Attainable By *S*'s Partner (Condition 1)	Prize Not Attainable By *S*'s Partner (Condition 2)	Prize Attainable By *S*'s Partner (Condition 3)	Prize Not Attainable By *S*'s Partner (Condition 4)
Block 2	58.33	57.55	59.00	21.67
Block 3	73.44	55.67	45.78	28.33
All trials	65.89	56.61	52.39	25.00

RESULTS

Dependency Differences Created by Differences in Goal Value

The *S*s in the first four conditions were led to believe that they and their partners were instrumentally interrelated in completing the assigned task. As is shown in Table 1, Conditions 1 through 4 had essentially similar means on the postsession questionnaire assessing the perceived instrumental relations between the team members. With perceived instrumentality constant, then, the direction and degree of dependency between team members in these conditions presumably varied directly with eligibility for the prize. Analyses of variance of the two sets of improvement scores in Conditions 1 through 4 revealed significant main effects for both *S Eligible for Prize* ($F = 15.35$; $p < .01$), and *S's Partner Eligible for Prize* ($F = 10.14$; $p < .01$). The first main effect attests to the success of the goal value manipulations. The more *S* stands to gain by completing the task rapidly, the faster he works. The second effect suggests that *S*'s group-task motivation increases when he perceives that other group members are highly dependent upon him for the attainment of their valued goals.

There were no significant differences in time scores for the baseline period among Conditions 1 through 4, so Table 2 presents only the mean improvement scores for the second and third trial blocks. With the exception of the differences between Condition 4 and the remaining conditions in both trial blocks, the effects of the perceived dependency relationships did not

become significant until Block 3. According to Hypothesis 2, *Ss* in Condition 1, who were told that both they and their partners were eligible for the valued rewards and, therefore, who may well have perceived a relatively high degree of interdependency, should be more highly motivated than the *Ss* in the other instrumentally interrelated conditions. This hypothesis was confirmed at an acceptable level of confidence ($p < .05$ by t tests) for each of the three comparisons with Condition 1 in the Block 3 data.

The only other significant difference in Block 3 was that between Conditions 3 and 4 ($p = .05$). In support of Hypothesis 3, *Ss* who were informed that the prize was attainable by their partners became highly motivated to perform their task even though they themselves were not eligible for the reward.

Effects of Perceived Instrumental Interdependency

Both *S-1* and his partner were eligible for the prize in Conditions 1 and 5, but the latter condition, it will be recalled, was designed to have low perceived instrumental interrelationships between the team members. Analysis of the perceived instrumentality scores indicated that the low instrumental dependency manipulation was not completely successful; as is shown in Table 1, the Condition 5 mean was not significantly lower than the Condition 1 mean ($p = .15$). A preliminary assessment of the motivational effects of this perceived instrumental interrelationship was obtained by discarding the two *Ss* with the lowest perceived instrumentality scores from Condition 1 and the two *Ss* with the highest scores from Condition 5 so that the conditions did differ significantly on this variable by t test ($p < .05$). The Block 3 difference in task motivation scores became significant with the elimination of these four *Ss* ($p < .05$), providing tentative support for Hypothesis 1. The improvement score means in Conditions 1 and 5 were now 81.00 and 49.57 respectively. Thus, *Ss* who believe they and their partners are instrumentally dependent upon each other for the attainment of a valued goal are more highly motivated than *Ss* who believe they and their partners can attain this goal independently of each other.

DISCUSSION

The obtained results generally support the hypotheses of the study. Perceived group member interdependency for reward attainment can increase an individual's motivation toward the assigned group task. The findings also suggest that this heightened motivation can be a cumulative effect reflecting the self-perception that he is dependent upon his partner and also the perception that the partner is dependent upon him.

If we assume that an individual who sees that the others in his group are dependent upon him and who also perceives that these others demand or expect certain task-relevant behaviors of him (in the present case, rapid work), the following general proposition can be formulated: *the strength of the motivation to conform to the task-relevant expectations of others in the group is a positive function of the perceived degree of interdependence among the members.* Intragroup opinion conformity under certain conditions, as well as many instances of behavioral conformity to group norms, can then be seen as special cases of this general proposition. For example, it is clear that pressures toward uniformity of opinion are heightened when "group locomotion" requires unanimity (6). To illustrate, Deutsch and Gerard (4) recently created this type of requirement in some of their experimental conditions by offering prizes to the best groups. The *Ss* given this instruction probably realized that any group member's chances of obtaining the desired goal was dependent upon the activity (i.e., expressed opinions) of the others in the group. Different instructions produced low perceived interdependency in other groups. As a result, the *Ss* in the highly interdependent groups exhibited a greater tendency to conform to the opinions of others in the group than the *Ss* in less interdependent groups even though it was fairly obvious that the group opinion was wrong and they could express their own opinions anonymously.

Thus, perceived high interdependence apparently produces conformity to group opinions as well as high group task motivation, supporting the hypothesis of a general relationship between perceived task interdependency and conformity to the task expectations of the others in the group. Attitudinal variables that

affect the degree to which interdependent members adopt each other's opinions, then, may generally affect behavioral conformity to relevant shared expectations in interdependent groups. Dittes and Kelley (5), employing interdependent groups faced with the requirement of consensus, found that the members' adoption of the group opinion was a function of the extent to which they felt accepted by the group. If the present general proposition is correct, differences in acceptance by the group also should produce differences in the relationship between perceived interdependence and group task motivation.

SUMMARY

The present study provides a direct test of an hypothesis invoked to account for the results of two preceding experiments, viz.: group members perceiving themselves as interdependent for their reward attainment become highly motivated toward the group task. An analysis also is made of the component relationships involved in perceived interdependence between members of a two-person group. The results generally are in accord with the following predictions: 1. Members of groups who are instrumentally interdependent become more highly motivated toward the group task than *S*s who can achieve their goals independently of the others in the group. 2. Under conditions of instrumental interdependence, *S*s who are told that both they and their teammates can attain a valued goal through the group activity become more highly task motivated than *S*s who are told either that only *S* or only *S*'s partner can attain this goal. 3. Given that the group members are instrumentally interdependent and also in a position to see each other's behavior, *S*s told that only their partners have access to a valued goal become more highly motivated toward the task than *S*s working for a goal that is less highly valued by both themselves and their partners.

REFERENCES

1. BERKOWITZ, L., & LEVY, B. I. Pride in group performance and group-task motivation. *J. abnorm. soc. Psychol.* 1956, **53,** 300–306.
2. BERKOWITZ, L., LEVY, B. I., & HARVEY, A. R. Effects of performance evaluations on group integration and motivation. *Hum. Relat.*, in press.
3. DEUTSCH, M. The effects of cooperation and competition upon group process. *Hum. Relat.* 1949, **2,** 199–231.
4. DEUTSCH, M., & GERARD, H. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51,** 629–636.
5. DITTES, J., & KELLEY, H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53,** 100–107.
6. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57,** 271–282.
7. PEAK, HELEN. Attitude and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955.* Lincoln, Neb.: Univer. of Nebraska Press, 1955. Pp. 148–188.
8. PEPITONE, EMMY A. *Responsibility to the group and and its effects on the performance of members.* Unpublished doctoral dissertation, Univer. of Michigan, 1952.

Received February 3, 1957.

THE EFFECT OF TRAINING UPON ACCURACY AND VARIABILITY IN INTERPERSONAL PERCEPTION[1]

W. J. CROW[2]

Behavior Research Laboratory, University of Colorado

One purpose of many training programs in clinical, counseling, and personnel psychology is to increase the accuracy of the trainee's perceptions of others. The ability to "understand others," or to have "clinical intuition" is considered a basic requirement for the good practitioner. It has been difficult to evaluate the success of such training, however, because until recently there has been no way to measure, by objective tests, the subtle and complex psychological variables that are presumably detected by the practitioner. Therefore, determination of improvement in clinical judgment has itself depended upon other clinical judgments. That is, either the accuracy of the students' perceptions would be directly rated by observers, or the students' judgments would be compared with the judgments of experienced practitioners.

Recent methodological improvements in the measurement of interpersonal perceptiveness have made it possible to measure the accuracy of clinical judgment in a more objective way. It is now customary to ask a person (the object) to fill out a personality questionnaire and then ask the trainee (the subject) to fill out the identical questionnaire just as he believed the object has filled it out. The measure of the subject's accuracy is the degree of correspondence between the object's actual responses and those which the subject estimated that the object had made. Such a technique has been used by Kelly and Fiske (7), for example, to measure the diagnostic competence of clinical trainees.

Early studies using this technique for measuring interpersonal perception focused primarily upon accuracy, but attention has been called recently to the importance of other components of the interpersonal-perception situation (1, 2, 5). Reasoning mathematically from an analysis-of-variance model, Cronbach (1) has advanced a hypothesis for the relationship between accuracy and one of these components—the amount of variability in the subjects' estimations:

> There is an optimal degree of differentiation in making judgments. If a judge [subject] can make accurate judgments as to the relative location of others [objects] on a continuum, then he is wise to make σ_y (the variability in his estimations) as large as σ_x (the variability in the others' scores)—never larger. But if he is forced to base his judgment on inadequate cues or if the available personality theory and situational knowledge do not permit trustworthy inference, then he should treat people as if they were very nearly alike. The person who attempts to differentiate individuals on inadequate data introduces error even when the inferences have validity greater than chance (1, p. 181).

It follows from Cronbach's hypothesis that, if the subject increases the variability of his judgments from one time to another and his ability to make accurate judgments does not increase correspondingly, then he will make greater errors the second time than the first. It also follows that whenever a judgmental task is very difficult or the subjects are very low in ability, subjects with little variability in their estimations will make small errors, and subjects with high variability will make large errors. Stated otherwise, accuracy and variability will be negatively correlated when the task is difficult or the subject is very inaccurate.

The purpose of the study herein reported was twofold: (*a*) to investigate the effects of a training program upon interpersonal perception, and (*b*) to test the hypothesis of the relationship between accuracy and variability in estimations.

To accomplish these purposes, interpersonal perception measures were administered at the beginning, during, and after the Senior year to a class of medical students who were divided into experimental and control groups. The experimental group was instructed in physician-patient relationships and was provided with the opportunity to establish such relationships through more prolonged contact

[1] This investigation was carried out as part of a larger study of medical education supported by the Commonwealth Fund of New York. The author is indebted to Dr. Kenneth R. Hammond for assistance and advice and to Dr. Fred Kern, Jr., for his assistance during many stages of the study.

[2] Now at San Diego College.

with patients than was possible for the control group. It was expected (*a*) that the students in the experimental group would become more accurate in interpersonal perception than students in the control group, and (*b*) that variability in estimation would be negatively related to accuracy of judgment.

METHOD

Subjects and Treatments

The *Ss* were 72 Senior medical students who were assigned at random to experimental and control groups. Approximately one fourth of the students were assigned to the experimental clinic during the first half of the year, and the remaining three-fourths were assigned to one of two surgery clinics or to an internal-medicine clinic. During the second half of the year, one of the first-half surgery groups was assigned to the experimental clinic, the other was assigned to the internal-medicine clinic, and the first-half experimental and internal-medicine groups were assigned to the surgery clinics. All students were tested at the beginning (Pretest), at the midpoint (Posttest I), and at the end (Posttest II) of their Senior year.

Measures

Thirty-six patients were selected at random[3] (with certain restrictions) from the medical wards of two hospitals, and each was interviewed by a physician. Six-minute sound films were taken of each of these interviews, and three randomly-selected sets of ten films were shown to the *Ss*, one of them at each of the three test administrations.

The *Ss* were required to perform two tasks: Task A was to estimate the self-ratings of the patients on seven personality scales, which were paraphrases of selected Minnesota Multiphasic Inventory scales (6); Task B was to estimate where the patient "really was" on the same set of scales. The patient's self-ratings were the criteria of accuracy on Task A; the patients' performance on the MMPI for Task B.[4]

Scoring

The Random-Comparison Method (2) was used to compute the scores, since this method yields a Differential-Accuracy score by eliminating artifactual components from the raw scores. In addition to the Accuracy scores, two response-set measures, which had previously been found to be stable aspects of the students' performance (4), were investigated in the present study: 1. *A measure of adherence to stereotype*, defined as the inverse of the Average-Variance score of an *S*'s estimations for the scales. (An *S* may perceive wide differences among people in the characteristic under consideration, or he may perceive people as being very much alike. He may, in other words, cluster his estimations closely or he may vary them widely. If an *S* clusters his estimations closely, his Average-Variance score will be small and his adherence to stereotype will be high. The variance of the estimations for each scale was computed and averaged for all scales as a measure of perceived individual differences.) 2. *A measure of assumed veridicality*, defined as the inverse of the Average Difference between the *S*'s estimation for the object's self-rating and the *S*'s estimation for where the object "really was." (An *S* could assume a large measure of agreement between an object's self-rating and "reality," or he could assume little agreement. If an *S* assumed a large measure of agreement, the Average-Difference score between his estimation for the object's self-rating and his estimation for where the object "really was" would be small, and the assumed veridicality would be high.)

The *Ss*' accuracy, adherence to stereotype, and assumed veridicality were measured in this fashion for each of the three administrations. Change scores were computed from pre- to posttest in order to compare the experimental and control groups and to permit investigation of the relationship between accuracy and the two response-set characteristics.

Effect of Training upon Interpersonal Perception

Table 1 presents the means and standard deviations of the change in the *Ss*' performance between the Pretest and Posttest I (first-half change) and between the Posttest I and Posttest II (second-half change) for the Accuracy, Average-Variance, and Average-Difference scores. The significance of the difference between the mean change for the experimental and control groups was computed by use of the *t* test for small samples.

Examination of the accuracy results revealed that the control-group improvement was significantly greater than the experimental-group improvement for one of the four measures (Task B, first half). The control-group improvement was also greater on two of the three remaining Accuracy scores, but the difference was not statistically significant. Contrary to what had been expected, the experimental group, which received training in interpersonal relations, did not improve more than the control group, which had not received such training. In fact, the trend of results suggests the opposite conclusion: that the training in interpersonal relations decreased accuracy.

Table 1 also presents the results for the Average-Variance scores. For each of the four measures the experimental-group change was significantly different from the control-group change. The experimental group in-

[3] For a discussion of importance of the manner of the selection of the objects in interpersonal perception studies see (3).

[4] For a detailed description of the measures and the scoring techniques see (2).

creased the average variance of their estimations more than did the control group from pre- to posttest. They exhibited more differentiation than did the control group; that is, they manifested less adherence to stereotype.

The results for the Average-Difference scores were similar. The experimental group changed more than did the control group in the amount of average difference that they saw between a patient's self-rating and where the patient "really was" on the personality scales. The results were statistically significant for both the first and second halves. Since the *S*s in the experimental group increased their Average-Difference scores, they assumed less veridicality than did the control-group students.

These results indicate that the effect of the training program for the experimental group was to increase the amount of differentiation that the *S*s made in their judgments about patients. There also appears to be a loss of accuracy, but the results were not conclusive.

Relationship between Accuracy and Variability in Estimations

The results obtained for the effect of training upon interpersonal perception were in accord with Cronbach's hypothesis that if an *S* increases his differentiation among objects without a corresponding increase in estimating ability, he will make greater errors. But the findings reported apply to the groups of *S*s only. A more direct way to study the relationship of variability and accuracy is to correlate the scores of the individual *S*s on the two measures; therefore, change in Accuracy scores was correlated with the change in in Average-Variance and Average-Difference scores for the first and second halves.

Table 2 presents the correlation coefficients obtained between *change* in variability and *change* in accuracy. Four of the eight correlations (all for Task B) were negative and significant beyond the .05 level of confidence. These results also support Cronbach's hypothesis.

TABLE 1

MEAN INCREASE, STANDARD DEVIATIONS AND *t* FOR DIFFERENCES BETWEEN GROUPS FOR INTERPERSONAL PERCEPTION SCORES

Measures	Experimental: 1st Half $N = 20$, 2nd Half $N = 17$		Control: 1st Half $N = 52$, 2nd Half $N = 55$		
	Mean Increase	*SD*	Mean Increase	*SD*	*t*
Accuracy scores					
Task A First half	1.70	5.95	3.88	5.92	1.36
Second half	−3.06	6.74	−3.89	6.17	0.46
Task B First half	−12.65	8.67	−6.81	2.78	2.46*
Second half	12.29	6.52	14.82	7.95	1.16
Average-Variance scores					
Task A First half	0.36	2.99	−1.51	3.43	2.08*
Second half	1.53	3.51	−0.29	2.50	2.30*
Task B First half	1.09	3.29	−1.57	3.47	2.85**
Second half	2.78	2.82	0.83	2.57	2.59*
Average-Difference scores					
First half	0.17	0.19	−0.06	0.26	3.48**
Second half	0.11	0.21	−0.06	0.21	2.78**

* $p = <.05$.
** $p = <.01$.

TABLE 2

CORRELATION COEFFICIENTS BETWEEN INCREASE IN ACCURACY, INCREASE IN AVERAGE-VARIANCE SCORES AND INCREASE IN AVERAGE-DIFFERENCE SCORES[a]

	Increase in Accuracy			
	Task A		Task B	
	1st Half	2nd Half	1st Half	2nd Half
Increase in Average-Variance scores				
Task A First half	−.19			
Second half		.13		
Task B First half			−.50*	
Second half				−.27*
Increase in Average-Difference scores				
First half	−.03		−.45*	
Second half		.07		−.26*

[a] $N = 72$.
* Significant at .05 level.

TABLE 3

CORRELATION COEFFICIENTS BETWEEN ACCURACY AND AVERAGE-VARIANCE AND AVERAGE-DIFFERENCE SCORES FOR THREE ADMINISTRATIONS

	Accuracy	
	Task A	Task B
Average-Variance		
Pretest	.51	.57
Posttest I	.30	.76
Posttest II	.37	.55
Average-Difference		
Pretest	.34	.57
Posttest I	.35	.56
Posttest II	.33	.49

$N = 72$.
$r = .28$ at the .01 level.

Table 3 presents the correlations for each administration betweer Accuracy scores and both Average-Variance and Average-Difference scores. All of the twelve correlation coefficients were negative and significant beyond the .01 level of confidence. As the *S*s made more variable estimations, their accuracy decreased. These results support quite conclusively the hypothesis that accuracy is negatively related to the amount of variability in *S*s' estimations.

DISCUSSION

Contrary to expectation, the group which received more training in interpersonal relations (the experimental group) did not improve in accuracy of interpersonal perception more than did the group without such special training (control group). In fact, the results support the opposite conclusion: that, as a result of their special training, the experimental group became less accurate than the control group.

The experimental group increased the variability in their estimations significantly more than did the control group. The experimental group also assumed significantly less agreement between the way *S*s would rate themselves and the way they "really were." These results, in conjunction with the fact that the experimental group decreased in accuracy, support Cronbach's hypothesis that, if an *S* increases his differentiation without a corresponding increase in ability, he will make greater errors. This hypothesis was further supported by the finding that individual Accuracy scores were negatively correlated with both Average-Variance and Average-Difference scores. Additional (though less conclusive) support for the hypothesis stemmed from the finding that *change* in variability from pre- to posttest was negatively correlated with *change* in accuracy. The more that *S*s increased the amount of variability they placed in their estimations, the less accurate they became. These results, taken together, constitute compelling evidence that the effect of the training program was to increase the variability that the trainees placed in their judgments and that their accuracy therefore decreased.

Investigation of the relationship between variability and accuracy in interpersonal perception has not progressed to the point where Cronbach's hypothesis can be recommended as a basis for action. However, the implications of the hypothesis are so important that they deserve statement, particularly in regard to training programs such as the one reported in this study. These results indicate that training programs devoted to increasing accuracy of interpersonal perception (as in medical education and in the training of teachers, clinical psychologists, and similar specialists) run the risk of decreasing accuracy when they increase the trainees' responsiveness to individual differences. Since very little is known about how to train people to make more accurate judgments about others, training programs frequently utilize a procedure of "exposure" and little else. The belief that placing the trainee in a position to observe others and to make judgments will produce desirable results is challenged by these findings. Such experience may lead the trainee to differentiate among people far beyond his capacity to do so accurately. In the absence of dependable measures of his accuracy, the trainee lacks knowledge of his errors and may continue inappropriate overdifferentiation long after the training has ceased.

Cronbach has discussed this problem as follows: "Recognizing an optimum degree of differentiation makes it necessary to re-examine and qualify statements commonly made in training teachers, to the effect that every pupil has his own pattern and the teacher must fit methods to that pattern, not treat the pupil in terms of the statistical average" (1, p. 183). Whatever the ethical, moral, or ideological justification for emphasizing the unique characteristics of each person's individuality, the evidence so far indicates that "differentiation is harmful if the extent of adaptation or differentiation exceeds the amount justified by the accuracy of social perception" (1, p. 183).

SUMMARY

The present study investigated the relationship between accuracy and variability in estimations and the effects of a training program in interpersonal relations upon interpersonal perception. Seventy-two Senior medical students were divided into an experimental group who received training in physician-patient relationships and a control

group who did not receive such training. At the beginning, during, and at the end of their Senior year, the students estimated the real status and self-ratings of patients presented in sound-film recorded interviews. Actual self-ratings and relevant MMPI scores were available for these patients as criteria.

Contrary to expectation, the experimental group tended to become less accurate than the control group. The experimental group increased the variability in their estimations significantly more than did the control group and also assumed significantly less agreement between the way subjects would rate themselves and the way they "really were." Accuracy scores were negatively correlated with variability scores and change in variability of estimation was negatively correlated with change in accuracy. These results, which agree with a hypothesis advanced by Cronbach, suggest that training programs devoted to increasing accuracy of interpersonal perception run the risk of decreasing accuracy when they increase the trainee's responsiveness to individual differences.

REFERENCES

1. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
2. CROW, W. J. A methodological study of social perceptiveness. Unpublished doctoral dissertation, Univer. of Colorado, 1954.
3. CROW, W. J. The need for representative design in studies of interpersonal perception. *J. consult. Psychol.*, 1957, in press.
4. CROW, W. J., & HAMMOND, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, in press.
5. GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.
6. HATHAWAY, S. R., & MEEHL, P. E. An atlas for the clinical use of the MMPI. Minneapolis: Univer. of Minnesota Press, 1951.
7. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951.

Received March 15, 1957.

REFERENCE GROUPS, MEMBERSHIP GROUPS, AND ATTITUDE CHANGE[1]

ALBERTA ENGVALL SIEGEL[2]
Stanford University
AND SIDNEY SIEGEL[2]
Center for Advanced Study in the Behavioral Sciences

IN SOCIAL psychological theory, it has long been recognized that an individual's *membership groups* have an important influence on the values and attitudes he holds. More recently, attention has also been given to the influence of his *reference groups:* the groups in which he aspires to attain or maintain membership. In a given area, membership groups and reference groups may or may not be identical. They are identical when the person aspires to *maintain* membership in the group of which he is a part; they are disparate when the group in which the individual aspires to *attain* membership is one in which he is not a member. It has been widely asserted that both membership and reference groups affect the attitudes held by the individual (4).

The present study is an examination of the attitude changes which occur over time when reference groups and membership groups are identical and when they are disparate. The study takes advantage of a field experiment which occurred in the social context of the lives of the subjects, concerning events considered vital by them. The subjects were not aware that their membership and reference groups were of research interest; in fact, they did not know that the relevant information about these was available to the investigators.

The field experiment permitted a test of the general hypothesis that both the amount and the direction of a person's attitude change over time depends on the attitude norms of his membership group (whether or not that group is chosen by him) and on the attitude norms of his reference group.

[1] This study was supported by grants from the Committee for the Study of American Values at Stanford University and from the Stanford Value Theory Project. We wish to acknowledge with gratitude the assistance given by Davis W. Thompkins, Marilyn Sanchez-Corea, and Coleen Baker in the execution of this study, and the generous administrative cooperation of Elva Fay Brown, Dean of Women at Stanford University, and her staff.

[2] On leave from Pennsylvania State University.

This hypothesis is tested with subjects who shared a common reference group at the time of the initial assessment of attitudes. They were then randomly assigned to alternative membership groups, some being assigned to the chosen group and others to a nonchosen group. Attitudes were reassessed after a year of experience in these alternative membership groups with divergent attitude norms. During the course of the year, some subjects came to take the imposed (initially nonpreferred) membership group as their reference group. Attitude change after the year was examined in terms of the membership group and reference group identifications of the subjects at that time.

THE FIELD EXPERIMENT

The *S*s of this study were women students at a large private coeducational university. The study was initiated shortly before the end of their freshman year, when they all lived in the same large freshman dormitory to which they had been assigned upon entering the university. At this university, all women move to new housing for their sophomore year. Several types of housing are available to them: a large dormitory, a medium-sized dormitory, several very small houses which share common dining facilities, and a number of former sorority houses which have been operated by the university since sororities were banished from the campus. These latter are located among the fraternity houses on Fraternity Row, and are therefore known as "Row houses." Although the Row houses are lower in physical comfort than most of the other residences for women, students consider them higher in social status. This observation was confirmed by a poll of students (5, p. 205), in which over 90 per cent of the respondents stated that Row houses for women were higher in social status than non-Row houses, the remaining few disclaiming any information concerning status differences among women's residences.

In the Spring of each year, a "drawing" is held for housing for the subsequent year. All freshmen must participate in this drawing, and any other student who wishes to change her residence may participate. It is conducted by the office of the Dean of Women, in cooperation with woman student leaders. Any participant's ballot is understood to be secret. The woman uses the ballot to rank the houses in the order of her preference. After submitting this ballot, she draws a number from the hopper. The rank of that number determines the likelihood that her preference will be satisfied.

In research reported earlier (5), a random sample was drawn from the population of freshman women at this university, several tests were administered to the *S*s in that sample, and (unknown to the *S*s) their housing preferences for the forthcoming sophomore year were observed by the investigator. The *S*s were characterized as "high status oriented" if they listed a Row house as their first choice, and were characterized as "low status oriented" if they listed a non-Row house as their first choice. The hypothesis under test, drawn from reference group theory and from theoretical formulations concerning authoritarianism, was that high status orientation is a correlate of authoritarianism. The hypothesis was confirmed: freshman women who listed a Row house as their first choice for residence scored significantly higher on the average in authoritarianism, as measured by the E-F scale (1, 2) than did women who listed a non-Row house as their first choice. The present study is a continuation of the one described, and uses as its *S*s only those members of the original sample who were "high status oriented," i.e., preferred to live in a Row house for the sophomore year. In the initial study (5), of the 95 *S*s whose housing choices were listed, 30 were "high status oriented," i.e., demonstrated that the Row was their reference group by giving a Row house as their first choice in the drawing. Of this group, 28 were available to serve as *S*s for the follow-up or "change" study which is the topic of the present paper. These women form a homogeneous subsample in that at the conclusion of their freshman year they shared a common membership group (the freshman dormitory) and a common reference group (the Row). These *S*s, however, had divergent experiences during their sophomore year: nine were Row residents during that year (having drawn sufficiently small numbers in the housing drawing to enable them to be assigned to the group of their choice) and the other 19 lived in non-Row houses during that year (having drawn numbers too large to enable them to be assigned to the housing group of their choice).

E-F scores were obtained from each of the 28 *S*s in the course of a large-scale testing program administered to most of the women students at the university. Anonymity was guaranteed to the *S*s, but a coding procedure permitted the investigators to identify each respondent and thereby to isolate the *S*s and compare each *S*'s second E-F score with her first.

To prevent the *S*s from knowing that they were participating in a follow-up study, several procedures were utilized: (*a*) many persons who had not served in the earlier study were included in the second sample, (*b*) the testing was introduced as being part of a nation-wide study to establish norms, (*c*) the test administrators were different persons from those who had administered the initial tests, (*d*) *S*s who informed the test administrator that they had already taken the "Public Opinion Questionnaire" (E-F scale) were casually told that this did not disqualify them from participating in the current study.

The *S*s had no hint that the research was in any way related to their housing arrangements. Testing was conducted in classrooms as well as in residences, and all procedures and instructions were specifically designed to avoid any arousal of the salience of the housing groups in the frame of reference of the research.

The annual housing drawing was conducted three weeks after the sophomore-year testing, and, as usual, each woman's housing ballot was understood to be secret. In this drawing, each *S* had the opportunity to change her membership group, although a residence move is not required at the end of the sophomore year as it is at the end of the freshman year. If an *S* participated in this drawing, the house which she listed as her first choice on the ballot was identified by the investigators as her reference group. If she did not, it was evident that the house in which she was currently a member was the one in which she chose to continue to live, i.e., was her reference group. With the information on each *S*'s residence choice at the end of

her freshman year, her assigned residence for her sophomore year, and her residence choice at the end of her sophomore year, it was possible to classify the subjects in three categories:

A. Women ($n = 9$) who had gained assignment to live on the Row during their sophomore year and who did not attempt to draw out of the Row at the end of that year;
B. Women ($n = 11$) who had not gained assignment to a Row house for the sophomore year and who drew for a Row house again after living in a non-Row house during the sophomore year; and
C. Women ($n = 8$) who had not gained assignment to a Row house for the sophomore year, and who chose to remain in a non-Row house after living in one during the sophomore year.

For all three groups of *S*s, as we have pointed out, membership group (freshman dormitory) and reference group (Row house) were common at the end of the freshman year. For Group A, membership and reference groups were identical throughout the sophomore year. For Group B, membership and reference groups were disparate throughout the sophomore year. For Group C, membership and reference groups were initially disparate during the sophomore year but became identical because of a change in reference groups.

As will be demonstrated, the Row and the non-Row social groups differ in attitude norms, with Row residents being generally more authoritarian than non-Row residents. From social psychological theory concerning the influence of group norms on individuals' attitudes, it would be predicted that the different group identifications during the sophomore year of the three groups of *S*s would result in differential attitude change. Those who gained admittance to a Row house for the sophomore year (Group A) would be expected to show the least change in authoritarianism, for they spent that year in a social context which reinforced their initial attitudes. Group C *S*s would be expected to show the greatest change in authoritarianism, a change associated not only with their membership in a group (the non-Row group) which is typically low in authoritarianism, but also with their shift in reference groups, from Row to non-Row, i.e., from a group normatively higher in authoritarianism to a group normatively lower. The extent ot attitude change in the *S*s in Group B would be expected to be intermediate, due to the conflicting influences of the imposed membership group (non-Row) and of the unchanged reference group (Row). The research hypothesis, then, is that between the time of the freshman-year testing and the sophomore-year testing, the extent of change in authoritarianism will be least in Group A, greater in Group B, and greatest in Group C. That is, in extent of attitude change, Group A < Group B < Group C.

RESULTS

Group norms. From the data collected in the large-scale testing program, it was possible to determine the group norms for authoritarian attitudes among the Row and the non-Row women at the university. The E-F scale was administered to all available Row residents ($n = 303$) and to a random sample of residents of non-Row houses ($n = 101$). These *S*s were sophomores, juniors, and seniors. The mean E-F score of the Row women was 90, while the mean E-F score of the non-Row was 81. The E-F scores of the two groups were demonstrated to differ at the $p < .001$ level ($\chi^2 = 11.1$) by the median test (6, pp. 111–116), a nonparametric test, the data for which are shown in Table 1.

Attitude change. The central hypothesis of this study is that attitude change will occur differentially in Groups A, B, and C, and that it will occur in the direction which would be predicted from knowledge of the group norms among Row and non-Row residents in general. The 28 *S*s of this study had a mean E-F score of 102 at the end of their freshman year. The data reported above concerning authoritarianism norms for all women residing on campus would lead to the prediction that in general the *S*s would show a reduction in authoritarianism during the sophomore year but that this reduction would be differential in the three groups; from the knowledge that Row residents

TABLE 1
FREQUENCIES OF E-F SCORES ABOVE AND BELOW COMMON MEDIAN FOR ROW AND NON-ROW RESIDENTS

	Residents of Non-Row Houses	Residents of Row Houses	Total
Above Median	36	166	202
Below Median	65	137	202
Total	101	303	404

TABLE 2
FRESHMAN-YEAR AND SOPHOMORE-YEAR E-F SCORES OF SUBJECTS

Group	E-F Score: End of Freshman Year	E-F Score: End of Sophomore Year	Difference
A	108	125	−17
	70	78	−8
	106	107	−1
	92	92	0
	80	78	2
	104	102	2
	143	138	5
	110	92	18
	114	80	34
B	76	117	−41
	105	107	−2
	88	82	6
	109	97	12
	98	83	15
	112	94	18
	101	82	19
	114	93	21
	104	81	23
	116	91	25
	101	74	27
C	121	126	−5
	87	79	8
	105	95	10
	97	81	16
	96	78	18
	108	73	35
	114	77	37
	88	49	39

generally are higher in authoritarianism than non-Row residents, the prediction based on social group theory would be that Group A would show the smallest reduction in authoritarianism scores, Group B would show a larger reduction, and Group C would show the largest reduction. The data which permit a test of this hypothesis are given in Table 2. The Jonckheere test (3), a nonparametric *k*-sample test which tests the null hypothesis that the three groups are from the same population against the alternative hypothesis that they are from different populations which are ordered in a specified way, was used with these data. By that test, the hypothesis is confirmed at the $p < .025$ level.

DISCUSSION

Substantively, the present study provides experimental verification of certain assertions in social group theory, demonstrating that attitude change over time is related to the group identification of the person—both his membership group identification and his reference group identification. The hypothesis that extent of attitude change would be different in the three subgroups of *S*s, depending on their respective membership group and reference group identifications, is confirmed at the $p < .025$ level; in extent of change in authoritarianism, Group A < Group B < Group C, as predicted.

Another way of looking at the data may serve to highlight the influence of membership groups and reference groups. At the end of the freshman year, the *S*s in Groups A, B, and C shared the same membership group and the same reference group. During the sophomore year, the *S*s in Group A shared one membership group while those in Groups B and C together shared another. From membership group theory, it would be predicted that the extent of attitude change would be greater among the latter *S*s. This hypothesis is supported by the data (in Table 2): by the Mann-Whitney test (6, pp. 116–127), the change scores of these two sets of *S*s (Group A versus Groups B and C together) differ in the predicted direction at the $p < .025$ level. This finding illustrates the influence of *membership* groups on attitude change. On the other hand, at the conclusion of the sophomore year, the *S*s in Groups A and B shared a common reference group while those in Group C had come to share another. From reference group theory, it would be predicted that attitude change would be more extensive among the subjects who had changed reference groups (Group C) than among those who had not. This hypothesis is also supported by the data (in Table 2): by the Mann-Whitney test, the change scores of these two sets of *S*s (Groups A and B together versus Group C) differ in the predicted direction at the $p < .05$ level. This finding illustrates the influence of *reference* groups on attitude change. Any inference from this mode of analysis (as contrasted with the main analysis of the data, by the Jonckheere test) must be qualified because of the nonindependence of the data on which the two Mann-Whitney tests are made, but it is mentioned here to clarify the role which membership and reference groups play in influencing attitude change.

The findings may also contribute to our understanding of processes affecting attitude change. The imposition of a membership

group does have some effect on an individual's attitudes, even when the imposed group is not accepted by the individual as his reference group. This relationship is shown in the case of Group B. If the person comes to accept the imposed group as his reference group, as was the case with the *Ss* in Group C, then the change in his attitudes toward the level of the group norm is even more pronounced.

Methodologically, the study has certain features which may deserve brief mention. First, the study demonstrates that it is possible operationally to define the concept of reference group. The act of voting by secret ballot for the group in which one would like to live constitutes clear behavioral specification of one's reference group, and it is an act whose conceptual meaning can be so directly inferred that there is no problem of reliability of judgment in its categorization by the investigator. Second, the study demonstrates that a field study can be conducted which contains the critical feature of an experiment that is usually lacking in naturalistic situations: randomization. The determination of whether or not a woman student would be assigned to the living group of her choice was based on a random event: the size of the number she drew from the hopper. This fact satisfied the requirement that the treatment condition be randomized, and permitted sharper inferences than can usually be drawn from field studies. Third, the test behavior on which the conclusions of this study were based occurred in a context in which the salience of membership and reference groups was *not* aroused and in which no external sanctions from the relevant groups were operative. This feature of the design permitted the interpretation that the E-F scores represented the *Ss'* internalized attitudes (4, p. 218). Finally, the use of a paper-and-pencil measure of attitude and thus of attitude change, rather than the use of some more behavioral measure, is a deficiency of the present study. Moreover, the measure which was used suffers from a well-known circularity, based on the occurrence of pseudo-low scores (1, p. 771; 5, pp. 221–222).

SUMMARY

In the social context of the lives of the subjects, and in a natural social experiment which provided randomization of the relevant condition effects, the influence of both membership and reference groups on attitude change was assessed. All subjects shared a common reference group at the start of the period of the study. When divergent membership groups with disparate attitude norms were socially imposed on the basis of a random event, attitude change in the subjects over time was a function of the normative attitudes of both imposed membership groups and the individuals' reference groups. The greatest attitude change occurred in subjects who came to take the imposed, initially nonpreferred, membership group as their reference group.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. GOUGH, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 237–246.
3. JONCKHEERE, A. R. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.
4. SHERIF, M., & SHERIF, CAROLYN W. *Groups in harmony and tension*. New York: Harper, 1953.
5. SIEGEL, S. Certain determinants and correlates of authoritarianism. *Genet. Psychol. Monogr.*, 1954, **49**, 187–229.
6. SIEGEL, S. *Nonparametric statistics: For the behavioral sciences*. New York: McGraw-Hill, 1956.

Received February 27, 1957.

SOCIOECONOMIC STATUS AND CONCEPTUAL THINKING[1]

JEROME SILLER

Prosthetic Devices Study, New York University

ALTHOUGH there is a considerable body of data indicating differences between socioeconomic groups on scores made on commonly used tests of intelligence, the meaning and significance of these differences is debatable. A better understanding of factors relating to socioeconomic group differences in mental abilities may be attained if we deal with separate phases of the intellective process rather than with intelligence in its more global and ambiguous sense. The present concern is therefore with the ability to conceptualize and with its related process, concept formation. For this purpose, the following characterization of concept formation is adequate: "Behavior governed by concepts requires that the same response shall be made to objects of the same class, but a different response, or no response at all to objects not belonging to that class" (21, p. 609).

Davis and Havighurst (2, 3) have pointed out that status differentiations have the effect in varying degrees of defining and limiting the developmental environment of the child. They, among others, maintain that social class is an important factor in determining the development of meanings.

Eells (6), a member of this "culture fair" group, purportedly demonstrated the existence of a middle-class bias on intelligence tests, after an analysis of individual test items. Without necessarily accepting Eells' conclusion about test bias, there are two aspects of his data which are pertinent for the present study. Eells reported that mean status differences are largest for verbal and smallest for nonverbal material, and that the quality of *abstractness* in the item appeared to produce large status differences. All differences were in the direction of high status superiority.

As in intelligence testing, there are virtually no definitive experimental studies relating conceptual ability to socioeconomic status. Two investigations concern themselves with physical causality (4, 10). Both report little relationship between the scores on their concept tests and socioeconomic status. Each study, however, has methodological limitations, i.e., Huang's study (10) has many uninterpretable responses that may have influenced the analysis, while Deutsche's work has been severely criticized by Lacey (12). Finally, both are based upon a restricted aspect of conceptualization.

Dixon (5) reports significant differences on a test of contradictory relationships between children two to six years old when compared on the basis of parental occupation (children of students and professors vs. working mothers). Dixon explains these differences in terms of differential intelligence.

Findlay and McGuire (7), using a variation of the Vigotsky test, hypothesized that if cultural bias affects test performance, children from lower social status backgrounds might be expected to exhibit a higher degree of problem solving capacity (whatever that may be) than children from higher social status backgrounds in order to achieve the same range of scores in mental test performance as the children with obvious cultural advantages. Children from HS and LS backgrounds (grade levels 3, 6, and 9) were equated on IQ as measured by a group intelligence test and compared on the supposedly less culturally weighted conceptual test. The hypothesis was not validated, and differences were found in the opposite direction.

The above studies have demonstrated that the content of the test item is not the only factor in social status group differences. Another often mentioned factor is the form of symbolism (e.g., verbal or nonverbal) by which the item is presented. Eells' (6) findings in this respect have already been mentioned. Tyler (19) demonstrated that when Eells' data were converted into standard score form, the mean status differences between verbal and nonverbal items were no longer significant. Havighurst and Janke (8) and McGurk

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in the Department of Psychology, Graduate School of Arts and Science, New York University. The author wishes to express his appreciation to Drs. J. V. Hanna, Marie Jahoda, Elsa E. Robinson, and S. A. Weiss. The valuable suggestions of Dr. Richard Christie throughout the study are deeply appreciated.

(16, 17) also report negative evidence in this respect. These studies indicate that it is by no means certain that class differences are more marked on verbal than nonverbal tests.

Conceptual ability in the abstract can only be expressed indirectly and through the media of certain forms of symbolism (verbal and nonverbal). The resultant group differences may not be differences in ability to conceptualize, as much as differences in proficiency (learned) with the type of symbolism used. The role of the supposed middle-class verbal superiority must thus be assessed. The initial purpose of this study was to determine whether status differences on conceptual tests are a function of the form of symbolism by means of which conceptual ability is being tested.

It has been noted that low status (LS) children tend to operate, as a group, on a more concrete level than do high status (HS) children (2, 14). Observation of the time perspective of LS and HS children indicates that the former are more restricted in this respect (13). Time perspective and sense of abstractness should be closely related, since a broad time perspective, by definition, corresponds with an abstract mode of response and a narrow time perspective with an immediate or concrete mode. This observation suggests that social status differences on tests of cognitive ability might also be a function of *actual* cognitive differences.

The present study therefore attempts an evaluation of the importance of the variables of form of symbolism and abstractness-concreteness for status differences on tests of conceptual ability.

The following hypotheses have been proposed as generalizations from findings in the area of intelligence testing to concept formation, and are tested in the present investigation.

Hypothesis 1. HS children score more highly than LS children on all tests of conceptual ability.

Hypothesis 2. The form of symbolism, verbal or nonverbal, of the conceptual test is a factor contributing to the status differences. Greater status differences are predicted on verbal material than on nonverbal material.

Hypothesis 3. When alternative answers are available which are alike in correctness but differ along the concrete-abstract dimension, HS children more frequently select abstract answers than do LS children.

The foregoing hypotheses pertain to direct comparisons of the two groups wherein the subjects are *representative* of their particular status. The following hypotheses pertain to comparisons in which subjects from each group are matched on the basis of a specified variable.

Hypothesis 4. When subjects from the two groups are matched on the basis of nonverbal scores, HS children score more highly than LS children on tests of verbal concepts dealing with the same kind of conceptualization as the nonverbal tests.

Hypotheses 5. When subjects from the two groups are matched on the basis of scores on verbal concept tests, LS children score more highly than HS children on tests of nonverbal concepts dealing with the same kind of conceptualization as the verbal tests.

Hypotheses 4 and 5 are directed, in part, toward the problem of the hierarchical development of concepts as discussed by Heidbreder (9). It is assumed in Hypothesis 4 that LS and HS children equated on nonverbal material will draw apart, to the advantage of the latter, when verbal material is introduced. An explanation of the status differences in terms of verbal facility (Hypothesis 2) would not be contributive to the issue of innate group differences vs. differential and unequal group experiences. The latter position would argue that LS individuals simply have not had the opportunity to achieve verbal skills. Advocates of innate group differences might advance an alternate explanation in terms of hierarchical development. They might argue that verbal superiority is a natural outgrowth of greater intellectual ability in that verbal material is less "thing-like" than nonverbal material and therefore requires a higher order intellectual process.

Hypothesis 5 is a more direct check on the "hierarchy" principle. If one assumes, as did Findlay and McGuire (7), that greater overall ability is required by the LS child to do as well as the HS child on tests of verbal concepts, then, when nonverbal material is introduced, the LS child should be at an advantage.

Validation of the two hypotheses would serve to rule out hierarchical order of concept attainment as a factor in this study. Failure to

do so, however, would not serve as evidence for either position.

METHOD

Subjects

The *Ss* were 181 white sixth-grade New York City public school children whose average age was about eleven. They were divided into two groups, HS and LS. There were 99 in the HS group and 82 in the LS group. Only white middle-class children in a middle-class school and white lower-class children in a lower-class school were used. In all, ten classes from four schools were tested.

The average IQ of the two groups differed significantly at well beyond the .001 level. The intelligence test was the Otis QSMA, Form M. Summary data are presented in Table 1. Table 1 shows also that the HS group has a mean reading score of 8.03 on the Metropolitan Achievement Elementary Reading Test, a score two years greater than the mean score of the LS group, although the LS group is approximately at the reading level appropriate for its grade.

Determination of Status

Status was primarily determined by free lunch data and a questionnaire filled out by the parents. The questionnaire, a modification of the Warner ISC (20), obtained information about education and occupation of both parents, number of people in the apartment and dwelling area. The 181 *Ss* finally used were from a larger pool of 309. The others were eliminated because of undetermined status, non-white, etc.

Procedure

Each *S* was given the complete battery of tests, unless absent from one of the testing sessions. All of the tests were group pencil-and-paper examinations with generous time limits. None of the sessions exceeded sixty minutes. Each class was visited three times for testing.

Description of Tests[2]

Classification tests. Two such tests were given: one verbal (VC) and the other nonverbal (NVC). The NVC test was part of the Lorge-Thorndike Intelligence Tests-Level 3, Form A (14). Three objects are shown to the left of the page. The *S* must pick one of the five objects to the right which best goes with the first three. The standardized time of nine minutes was allowed for the 25 items.

TABLE 1
READING AND IQ CHARACTERISTICS OF SAMPLE

Variable	*N*	Mean	*SD*	Range	*t*
IQ					
High status	99	118.31	14.08	86–149	6.27**
Low status	78	102.65	17.82	69–142	
Reading					
High status	99	8.03	2.06	3.4–11.6	21.7**
Low status	76	5.97	1.97	3.0–11.6	

** Significant beyond the .01 level.

The VC test was specially constructed by the *E* from items reported in the literature and judged to be appropriate for the sixth grade. Attention was given to "culture fair" criticisms of item construction. The task of the *S* was to select the word that was most different from the remaining three. Ten minutes were allowed for the 35 items.

Analogies tests. Two analogies tests were given: one verbal (VA) and the other nonverbal (NVA). The NVA test was also part of the Lorge-Thorndike Intelligence Tests-Level 3, Form A (14). The ten minutes allowed for the 30 items was one minute more than the standard time allowance. The VA test was specially constructed by the *E* in a manner similar to the VC test. Testing time for the 35 items was 10′15″.

The scoring for the classification and analogies tests was simply the number of correct answers.

Abstract-Concrete Similarities (A-C Sim.) A number of recent studies have successfully revealed qualitative differences on vocabulary tests. The procedure, in multiple-choice tests, has been to have all of the choices correct but ranging in quality from concrete to abstract. The task of the *S* is to pick what he believes to be the best answer of the ones offered. This procedure has served to demonstrate developmental differences (11) as well as normal-schizophrenic differences (1). The present administration is an attempt to extend its use to a single age group with the independent variable being social status.

The work of Chodorkoff and Mussen (1) served as a model for a specially constructed, objective vocabulary test suitable for sixth graders. The test, consisting of 35 words, was

[2] Two additional tests were used in the original study (18): one was intended to be a nonverbal analogue of the Abstract-Concrete Similarities Test and the other a measure of status-oriented attitudes. The former proved to be technically inadequate and the latter is not relevant to the immediate problem. Discussion of these tests is therefore omitted here but may be found in the dissertation.

to determine group preferences in type of definition selected. For each word, definitions in terms of class, description, example, function, and incorrect choice were constructed. The order of definitions was nonsystematic. A sample item follows: *Dog:* (*a*) it has four feet, a tail, and fur (description); (*b*) it means a pet dog (example); (*c*) it's a small animal (class); (*d*) something wooden (incorrect); (*e*) it barks and guards your house (function). No time limit was set. Each protocol was given five scores corresponding to the number of times that each type of definition was used.

RESULTS

Status Comparison: Unmatched Groups

Hypothesis 1. Testing of the first hypothesis involved the NVA, NVC, VA, and VC tests. Status comparisons were made by means of the Median test. Distribution free methods were used because most of the distributions were significantly skewed and remained so despite attempts at transformations. The skewness appears to be a result of excessive time limits.

The HS children did better than the LS children on all of the tests of conceptual ability, three of the four tests showing significant differences. Hypothesis 1 was thus substantiated. Summary data for the status comparisons are presented in Table 2.

Hypothesis 2. The hypothesis and accompanying prediction were borne out. As can be noted in Table 2, there were clearly significant differences between the status groups on the verbal tests. Only the NVC of the two nonverbal tests was found to have a median difference between groups great enough to be statistically significant at the .05 level, although the NVA was close to significance.

The absolute point differences between the groups on the verbal tests were greater than the point differences on the nonverbal tests. Direct comparison, however, was not feasible as conversion to standard score units was not possible. However, findings discussed below in relation to the matched groups also support the hypothesis. To anticipate, no group differences were found when verbal scores were matched and comparisons made on nonverbal material, but, when the reverse was done, there were significant differences on the compared verbal material.

Hypothesis 3. Testing the third hypothesis involved the determination of which group selected more of each type of definition on the A-C Sim. test. Table 3 presents the group totals for each type of definition. Chi square was computed to test the hypothesis that choice of definition is independent of status. Since the chi square for 4 *df* was 102.05, significant at well beyond the .001 level, the hypothesis of independence could be rejected with confidence.

Examination of the contribution of each cell indicated that the class, example, and incorrect categories were solely responsible for the significant chi square. The actual difference, in terms of number of class responses, is small but statistically significant. There appears to be a greater proneness to use abstractions among the HS group than among the LS group.

A separate analysis of the importance of reading ability for choice of definition indicated that the two-year difference in group reading scores was not sufficient to explain the status differences. Proneness to respond in an abstract fashion was still more characteristic of the HS group than of the LS group even when the most inferior readers of the LS group were removed.

The significant difference should not obscure

TABLE 2

NONPARAMETRIC MEDIAN TESTS OF DIFFERENCES BETWEEN THE HS AND LS GROUPS ON THE NVC, NVA, VA, AND VC TESTS

Test	Status Group with Larger Score	p*
NVC	High	.025–.05
NVA	High	.05 –.10
VC	High	< .001
VA	High	.025–.05

* Median test values were corrected for continuity.

TABLE 3

COMPARISON OF THE LOW AND HIGH STATUS GROUPS IN PERCENTAGE SELECTING EACH TYPE OF DEFINITION

(Total Number of Responses: HS = 3311, LS = 2642. Percentages Derived for Each Group Separately.)

Status	Description	Class	Function	Example	Incorrect	Total
High	38.0	33.6	23.1	4.0	1.2	99.9
Low	36.4	27.9	23.6	9.2	3.0	100.1

the close similarity between the responses of the two groups. The type of choice is distributed in the same order with both groups having as an order of preference, description, class, function, example, and finally incorrect. It is questionable whether the group difference on abstraction is socially as well as statistically significant.

Status Comparisons: Matched Groups

Data bearing on Hypotheses 4 and 5 were obtained by matching members of each group on the basis of score on one form of symbolism and comparing them on the test of the same function using the other form of symbolism. Four such matchings were made using the classification and analogies tests. Analysis was by means of the Sign test, corrected for continuity.

Hypothesis 4. Two nonverbal matchings were made. NVC scores were equated and VC scores compared, and NVA scores equated and VA scores compared. Both matchings showed a significant difference in favor of the HS group. Hypothesis 4 was confirmed. The data are summarized in Table 4.

Hypothesis 5. Two verbal matchings were made. VC scores were equated and NVC scores compared, and VA scores were equated and NVA scores compared. The data, as shown in Table 4, indicate that the hypothesis was not confirmed by either of the matchings. It was not possible to assess the importance of order of attainment.

Excluded cases. The number of matchings for each test was always less than the total number of cases involved for each of the tests. Analysis of the equated groups indicated that they were not representative of their status levels. They were composed primarily of the lower and middle range of scores of the HS population and the middle and higher range of scores of the LS population. It was possible to equate many *Ss* in the upper range of scores; however, the HS *Ss* in this range could not be completely matched by the LS *Ss* since there were more of the former in the sample. LS and HS *Ss* could not be matched on the lower range of scores to any appreciable degree because there were very few HS *Ss* who scored very low on the various tests while there were many LS *Ss* who did so. The excluded *Ss* represent a one-tailed extreme for each of the status groups, the implications of which fact will be considered below.

TABLE 4

NONPARAMETRIC SIGN TESTS OF DIFFERENCES BETWEEN MATCHED GROUPS ON THE NVC, NVA, VA, AND VC TESTS

Test Matched	Test Compared	*N*	HS[a]	LS[a]	Ties	*p*[*]
NVC	VC	68	45	22	1	.007
NVA	VA	60	39	21	0	.028
VC	NVC	56	22	32	2	—
VA	NVA	55	28	27	0	—

[*] Sign test values were corrected for continuity.

[a] Number of times member of that status exceeds matched mate.

Status Comparisons Matched for Intelligence

Fifty *Ss* from each group were closely matched for intelligence. The mean IQ for the LS group was 111.96 and 112.00 for the HS group. All 28 of the excluded LS *Ss* had IQs below 100. All but one of the excluded HS *Ss* had IQs over 100. Again, we note that the equated groups were not unbiased samples of their parent population. There were no significant differences between the equated groups when comparisons by means of the Sign test were made on the NVC, VC, NVA, and VA tests. Unlike the findings of Findlay and McGuire (10) who used IQs ranging only from 90–110, the direction of differences did not favor either status.

DISCUSSION

So far, emphasis has been on defining the functional importance of the variables, form of symbolism and proneness to use abstractions, for a generalized consideration of conceptualization. This approach is in contrast to the few previous studies in the area which have generally considered restricted aspects of conceptualization.

Examination of the "one-tailed status extremes" excluded in matching facilitated insight into the meaning and possible basis of the status differences. The excluded LS *Ss*, it will be remembered, were from the low range of scores in the LS population and the excluded HS *Ss* from the high range of scores in the HS population. Only because of the apparently strong relationship between verbal performance and status were significant differences found for some of the equated

groups. On the other hand, when reading score was equated, as was done on the A-C Sim. test, and, in effect, only the middle range of *S*s compared, no over-all status differences were found in type of definition. Again, there were no significant differences when the groups were based on IQ. It was obvious from an inspection of the frequency distributions in all of these instances that important segments of each of the status groups (the one-tailed status extremes) and most particularly the LS, were not entering into the group comparisons.

A fraction of the LS group consistently scored low on most or all of the tests. Placement in the bottom quartile on three or more tests was arbitrarily fixed as the criterion for terming an *S* a "low scorer." There were 19 such individuals of the 82 in the LS group. There was no subgroup corresponding to the "low scorers" in the HS population although about a half-dozen individuals did score consistently low on more than two of the tests. However, even these individuals had scores which generally were higher than that of many of the LS "low scorers."

Having isolated this "low scorer" group it was possible to evaluate their role in the production of the differences between the HS and LS groups. Statistical comparison of the groups, with only the "low scorers" omitted, yield similar medians, distributions, etc. Comparison of the frequency distributions clearly indicates that the statistically significant group differences arose much more out of the infrequency of low scoring HS *S*s than from the lack of high scoring LS *S*s. For the most part, the two status samples are very similar in test performance, and the status differences are a function of a relatively small number of *S*s in the LS group.

On the basis of the above analyses, it appears to be misleading to refer to general status differences in conceptual ability, even though such differences can be statistically derived. A more accurate statement of the situation as it appears from the present study would be that when one considers status differences in conceptual ability there is a greater possibility of finding *S*s with low ability in an LS group than in an HS group but that, for the most part, performance of LS and HS *S*s is similar.

Why should there be a prevalence of "low scorers" in the LS sample and inferentially in the LS population? It is possible that an answer to this question may shed new light in the recurrent controversy over the source of status differences in cognitive ability. Although the study was not designed to provide such an answer, it does suggest a hypothesis which might contribute to an answer: that the "low scorers" might be distinguished from the other LS *S*s in that the former come from families that are not oriented in terms of upward social mobility while the latter come from upwardly mobile homes. As a concomitant of this attitude toward social mobility we might expect greater psychological impoverishment in the daily existence of "low scorers" than in the other LS *S*s. Status in the present study was defined in objective sociological terms and not as the families considered themselves. It is likely that *S*s from upwardly mobile LS families have psychological environments more similar to HS families than to sociologically determined LS families who are not so oriented.

Social immobility, of course, is not the exclusive or even necessarily the most important factor to consider in reference to the "low scorer" group. There unquestionably are some children in the group with inherently low intellectual endowment, as there are others who have been retarded by experiential impoverishment. Other factors to be evaluated are prenatal environment, nutritive deficiencies, and, in this particular sample, a greater proportion of first generation Americans in the LS than in the HS group.

It is obvious that the necessary data for explaining the basis for status differences will not be derived from the psychometric type of study (such as this one) which ends by showing differences, when an explanatory study should begin from this point. However, if other studies show effects similar to that of the "low scorer" group in influencing the results and if provision is made to analyze the characteristics of these *S*s and their families, the meaning of average status differences may be clarified.

SUMMARY

This study was designed to relate socioeconomic status to conceptual ability. The basic problem was to evaluate the importance of (*a*) the form of symbolism by means of

which conceptual ability was being tested and (*b*) the abstractness of the material for status differences in performance.

A measure of status attitudes and four conceptual tests were administered to groups of 99 high status and 82 low status white sixth-grade public school children.

The major results were as follows:

1. As groups, HS children did better than LS children on all tests of conceptual ability, particularly those involving verbal material.

2. The HS group selected more definitions of an abstract type than did the LS group.

3. The status groups were matched for a number of variables such as verbal score, nonverbal score, and intelligence, and then compared on a number of other variables. Generally there were no significant status differences. This procedure was used to determine whether the hierarchical order of concept attainment principle could be ruled out as a factor in HS verbal superiority. The results were inconclusive.

4. The matchings usually excluded low scoring *Ss* in the LS group and high scoring *Ss* from the HS group. When the relative small number of lowest scorers in the LS group was removed from the group comparisons there were no status differences. The meaning of general group status differences was questioned and further study of the "low scorer" group strongly suggested.

REFERENCES

1. CHODORKOFF, B., & MUSSEN, P. Qualitative aspects of the vocabulary responses of normals and schizophrenics. *J. consult. Psychol.*, 1952, **16,** 43–48.
2. DAVIS, A. *Social-class influences upon learning.* Cambridge, Mass.: Harvard Univer. Press, 1948.
3. DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child-rearing. *Amer. sociol. Rev.*, 1946, **11,** 698–710.
4. DEUTSCHE, JEAN M. The development of the children's concepts of causal relations. In Barker-Kounin-Wright (Eds.), *Child behavior and development.* New York: McGraw-Hill, 1943, Ch. VIII.
5. DIXON, J. C. Concept formation and emergence of contradictory relations. *J. exp. Psychol.*, 1949, **39,** 144–149.
6. EELLS, K., DAVIS, A., HAVIGHURST, R. J., HERRICK, V. E., & TYLER, R. *Intelligence and cultural differences.* Chicago: Univer. of Chicago Press, 1951.
7. FINDLAY, D. C., & MCGUIRE, C. Social status and abstract behavior. *J. abnorm. soc. Psychol.*, 1957, **54,** 135–137.
8. HAVIGHURST, R. J., & JANKE, L. L. Relations between ability and social status in a midwestern community: I. Ten-year-old children. *J. educ. Psychol.*, 1944, **35,** 357–368.
9. HEIDBREDER, EDNA. The attainment of concepts: I. Terminology and methodology. *J. gen. Psychol.*, 1946, **36,** 173–189.
10. HUANG, I. Children's conception of physical causality; a critical summary. *J. genet. Psychol.*, 1943, **63,** 71–121.
11. KRUGLOV, LORRAINE P. Qualitative differences in the vocabulary choices of children as revealed in a multiple choice test. *J. educ. Psychol.*, **44,** 1953, 229–243.
12. LACEY, J. I. Review of Deutsche (see 7). *Amer. J. Psychol.*, 1939, **52,** 318–320.
13. LESHAN, L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, **47,** 589–592.
14. LORGE, I., & THORNDIKE, R. L. *The Lorge-Thorndike Intelligence Tests.* Boston: Houghton Mifflin, 1954.
15. MCCANDLESS, B. R. Environment and intelligence. *Amer. J. ment. Def.*, 1952, **56,** 674–691.
16. MCGURK, F. C. J. Socioeconomic status and culturally weighted test scores of Negro subjects. *J. appl. Psychol.*, 1953, **37,** 276–277.
17. MCGURK, F. C. J. On White and Negro test performance and socioeconomic factors. *J. abnorm. soc. Psychol.*, 1953, **48,** 448–50.
18. SILLER, J. The effect of differential socioeconomic status upon the development of conceptual ability. Unpublished doctoral dissertation, New York Univer., 1956.
19. TYLER, F. T. Comments on the correlational analysis reported in intelligence and cultural differences. *J. educ. Psychol.*, 1953, **44,** 228–295.
20. WARNER, W. L., MEEKER, M., & EELLS, K. *Social class in America.* Chicago: Science Research, 1949.
21. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

Received March 4, 1957.

RECALL OF INTERRUPTED TASKS UNDER STRESS: A PHENOMENON OF MEMORY OR OF LEARNING?[1]

ALBERT J. CARON AND MICHAEL A. WALLACH
Harvard University

PERHAPS no experimental approach to repression has received more attention than the attempt to administer Zeigarnik's interrupted-task procedure (22) under ego-involving conditions. The rationale underlying this use of the Zeigarnik technique held that ego involvement would cause incompleted and completed tasks to be viewed as failures and successes respectively, and hence, in accordance with repression theory, would lead to the expulsion of incompleted items from consciousness. The prediction followed, therefore, that the typical "Zeigarnik" recall pattern (superior recall of incompleted tasks) would be replaced by a "repressive" recall pattern (superior recall of completed tasks).

Since this view was first advanced, much evidence has accumulated (1, 2, 3, 7, 8, 18, 19) which indicates, not surprisingly, that such a reversal of the Zeigarnik ratio is not a function of stress alone, but depends, in addition, on personality factors—some individuals recalling a preponderance of successes and others yielding extreme Zeigarnik ratios under ego-involving conditions. While the isolation of the particular personality dimensions associated with each of these stress recall patterns offers a fertile field for personality research, the mechanisms assumed to underlie each pattern —namely, repression of failures in the case of selective recall of successes (S-recall) and superior trace stabilization of failures in the case of selective recall of failures (F-recall)—are still open to question. As others (15, 20) have observed, finding a selective recall preference for success items under stress is not definitive proof that failure items were repressed. An alternative explanation, selective *learning* in favor of success items, has not been ruled out. Likewise, a selective recall preference for failure items under stress may arise not from enhanced stabilization of their traces but rather from their being better learned in the first place. The obvious experimental solution here—to equate successes and failures for original learning—is hard to achieve, for if failure items are really threatening, then the same factors which theoretically lead to their repression (or trace stabilization) may also be expected to impair (or improve) their learning.

The aim in the present study was to provide a more feasible test of the memory versus learning explanations of both these recall patterns. The approach employed follows from an analysis of the assumptions underlying interpretation of each pattern as a memory or learning effect. Consider first the alternative assumptions with respect to S-recall under stress.

As has been noted (4, 23), the essential postulate of repression theory is that threatening events are not actually forgotten but persist in an unconscious state, continuously striving to regain consciousness. Moreover, should the threatening character of repressed events be allayed (via psychotherapy or other means), they should *re-emerge* into consciousness. The selective learning position, on the other hand, implies no such restoration for forgotten items: it assumes that decreased recall results entirely from a deficiency in original registration. The two interpretations thus offer different predictions for the recall of events whose threatening nature is allayed after learning—the repression view implying enhanced recall, the selective learning view implying no change in recall, for such items. In the present study, two groups (both known to exhibit S-recall under stress) were subjected to a stressful interrupted-task procedure. For one group, however, recall was preceded by a quasi-therapeutic session in which the test was exposed as an experimental hoax. If the S-recall pattern is indeed due to repression, such exposure, by reducing associated anxiety, should restore the recall of incompleted items, and hence cause this group to shift toward a Zeigarnik pattern. If, on the other hand, selective learning underlies S-recall, then no such shift should occur.

With respect to the F-recall pattern, ex-

[1] This research was conducted while the senior author held a U.S.P.H.S. postdoctoral research fellowship at Harvard University. Grateful acknowledgment is made to Dr. Jerome S. Bruner for making available the facilities of his laboratory.

planation in terms of selective learning is much like that offered above for S-recall. Recall is said to be totally dependent on learning and, whatever the reason may be, failure items have registered more strongly than success items. The interpretation in terms of selective remembering, on the other hand, is based, according to Rosenzweig's recent formulation (17), on the task-tension theory of Lewin (12) and Zeigarnik (22). According to this view, the traces of failure items are better stabilized than those of success items, because the former are laid down in a region of high psychic tension—tension aroused by instructions to perform a task and dischargeable only by its completion. Rosenzweig has called this a "need-persistive" reaction to failure as contrasted with the "ego-defensive" reaction of repression. This position predicts that anything done after learning to reduce task tension should bring about a reduction in the recall of failure items. The selective learning view, on the other hand, implies no recall change as a consequence of such post-learning manipulation. To test these predictions, two other groups—both known to exhibit F-recall under stress—were subjected to the same stressful situation described above, and recall for one group again followed exposure of the experimental hoax. If the F-recall pattern is indeed due to superior trace stabilization of failures, the revelation that incompleted tasks were actually impossible to solve should serve to discharge any residual tension associated with incompletion and thereby tend to reduce the recall of failures. No such reduction in recall of failures should occur, on the other hand, if this effect is due to differential learning.

A final question examined in the present study may now be mentioned. If our results should indicate that both the S- and F-recall patterns are a function of learning rather than retention, does this mean that there is no evidence in the present experiment for the operation of mnemonic processes? This question is prompted by consideration of an aspect of the present kind of procedure which distinguishes it from the original Zeigarnik technique, a difference heretofore neglected by most writers. To construct a situation sufficiently threatening for college students, the experimental tasks have usually been presented in the guise of an intelligence test and indeed are so presented here. In so doing, however, the procedure is given a unity and integration which is not present in the typical Zeigarnik situation, where *S* is merely asked to perform a series of disparate and ostensibly unrelated tasks. Under intelligence-test conditions, then, it is entirely possible that *S*'s defensive reaction is oriented more toward the test as a whole, i.e., toward the single experience of examination failure, rather than toward particular items within the test, and hence selective recall of individual items may not be a measure optimally sensitive to *S*'s invoked defense. Memory for the entire test situation would seem more pertinent, but for obvious reasons such a measure is unfeasible. To determine whether there was indeed a mnemonic reaction to the total test situation in this study, an alternative indicator in the area of perception was employed. The rationale for its use is as follows: If *S* tends to repress an experience, aspects of that experience should be relatively inaccessible as hypotheses for perceptual recognition. Likewise, if the trace of an experience is well stabilized for *S*, aspects of it should be hyperaccessible as recognition hypotheses. Thus, following the stress situation, words related to it in a general way were presented tachistoscopically (along with matched neutral words). A repressive reaction to the entire situation should result in heightened recognition thresholds for such words; enhanced trace stabilization, on the other hand, in lowered thresholds. Moreover, if these effects are truly memorial, and not due to registration processes, they should cease to occur when exposure of the experimental hoax precedes the perceptual test.

METHOD

Subjects

Since none of the personality or performance measures previously shown to relate to selective recall under stress have been conclusively validated, it was not considered appropriate to use any one or combination of them as a basis for preselecting the S-recallers and F-recallers in our various treatment groups. As an alternative to preselection, it was decided to administer a large battery of promising measures[2] to randomly selected *S*s already tested in our experimental conditions, to factor analyze these measures, and, from the correlations between the resulting factors and the

[2] The measures used included tests related rationally as well as experimentally to each of three personality areas: achievement motivation, hysteria, and ego strength.

selective recall scores of our stress group, to isolate a stable predictor that could be used *post hoc* to subdivide all treatment groups. This analysis yielded four orthogonal factors, only one of which was at all related to recall direction in the stress condition. The measure with the highest loading in this factor (and, incidentally, with negligible loading in the other factors) was the dichotomous variable of private prep school versus public high school attendance, and indeed this single variable was more strongly associated with our recall measure[3] than any other variable or combination of variables defining this factor (see Footnote 4 for the major variables). Thus, the public-private dichotomy involved fewer dimensions than one might at first have supposed, and also accounted for most of the factorial relationship with recall direction. These considerations, coupled with considerable evidence suggesting that the reactions of Harvard public and private school *Ss* should be quite different in an intelligence-test situation,[4] led us to use this variable to subdivide our groups into S-recaller and F-recaller components.

The *Ss* were chosen in the following way: Invitations to participate in the experiment were mailed out to every tenth name in the list of Harvard freshmen. The list was gone through twice in this manner, starting at a different name each time. The letter attractively described the experiment as a study of "symbolic processes," and also emphasized the remuneration for participation. About half of the men to whom letters were mailed volunteered—125 in all, of whom 117 completed the experiment—and they were randomly assigned to the several experimental conditions. The number of public and private school *Ss* in each condition was roughly equivalent.

Sequence of Procedures

The order of procedures for all *Ss* was as follows: (*a*) initial digit recall test (4 minutes); (*b*) scrambled sentence tasks (50 minutes); (*c*) second digit recall test (4 minutes); (*d*) interim discussion with *E* (8–10 minutes); (*e*) initial recall of scrambled sentences (5 minutes); (*f*) determination of perceptual recognition thresholds (60 minutes); (*g*) second recall of scrambled sentences (5 minutes); (*h*) third recall of scrambled sentences, two days later (5 minutes).

Tests

Digit recall. Half of the Wechsler-Bellevue Digit Recall Test (21) was administered before, and half after, the scrambled sentence tasks, one instance of each forward and backward couplet being presented in each administration. All responses were written from left to right, and writing never began until *E* finished reading the digits. The number of digit sequences completely correct provided the metric. Scores for the first half alone constituted a check on the equality of memory ability for the several experimental groups; the difference between scores for the first and second halves provided a measure of the anxiety induced by the intervening work on scrambled sentences.[5]

Scrambled sentences. The experimental tasks were sixteen 20-word sentences similar to ones used by Alper (1), Eriksen (6), and Caron (5). Each sentence was divided into 8 to 10 two- and three-word phrases and was presented in scrambled form, *S*'s job being to rearrange them into a meaningful sentence. Half of the sentences, randomly distributed in the series of sixteen, were unsolvable. Contents of the sentences were diversified so that recall of one would be no aid to recall of another, and the completed and incompleted sentences were equated for recall difficulty in a previous study [see Caron (5)]. To facilitate group administration, the sentences were presented in booklet form, one sentence to a page. A 2¼-min. time limit was set for each solution and all *Ss* were required to start a new sentence at the same time. As an aid to recall, each sentence was named by selecting one phrase that best represented its meaning and placing this phrase above the scrambled sentence in the test booklet. *E* called out this "name" phrase at the start of each problem. For both initial and later recall tests, *S* had no prior knowledge that recall would be requested. In recalling sentences, *S* was required to write the name of a sentence or any other key phrase or word sufficient to identify it without question, no more than five minutes being allowed for each recall session. The measure of recall direction was the number of completed minus the number of incompleted tasks recalled (CR-IR).

Perceptual materials. Immediately after initial recall, perceptual recognition thresholds were determined for words pertaining in a general way to the scrambled sentence test and for neutral wo. of matched structure and of matched frequency (according to the Thorndike-Lorge L count). The former group included the following words: *sentence, booklet, phrases, examination, reasoning, scrambled,* and *tested;* the corresponding neutral words were *sunshine, cookies, sketches, agriculture,*

[3] Chi square was significant at less than the .025 level with a one-tail test—use of a one-tail test being justified since recall directions were predicted from the findings discussed in Footnote 4.

[4] McArthur's studies (13, 14) of the personalities of Harvard public and private school boys, for instance, indicate that the former tend to have a stronger drive for achievement and to be more sure of their personal and intellectual worth than the latter. Research by Atkinson (3) provides further support for this view. *Ss* with high need for achievement recalled more failures under stress and more successes under neutral conditions, while *Ss* with low need for achievement did the opposite—results which directly parallel our recall findings for public and private school *Ss* respectively. Finally, we may consider the performance of public and private school boys on some of our independent measures. The public school students made significantly higher grades at Harvard than those from private schools ($p < .02$ with a two-tail test), although the two groups did not differ in intelligence. Moreover, on the factor mentioned above, public school attendance was associated with such achievement variables as level of aspiration, while private school attendance was related to such indications of personal insecurity as the Knutson Personal Security Inventory (11) and the Social Introversion scale of the Guilford STDCR battery (10).

[5] For the use of digit recall as a measure of anxiety, see Rapaport *et al.* (16).

leisurely, channeled, and *regard*. *Ss*, tested in groups of four, were seated five feet from a ground glass screen in a room illuminated solely by a 40-watt fluorescent ceiling bulb one foot behind *S*. Each word, no more than 14 inches in length, was flashed at 1/100-second, beginning at subthreshold brightness levels and rising in two-volt steps until all *Ss* recognized the word. Threshold was taken as the voltage level at which correct report first occurred. The order of threatening and neutral words was random, with matched words always being seven positions apart. To prevent communication between *Ss*, responses were written, *E* observing these and telling each *S* to stop and turn over his paper when he had written the correct word.

Conditions

Stress. The stress condition was structured with two objectives in mind. (*a*) to convince *S* that he was taking a highly valid and discriminating intelligence test, and (*b*) to make him feel that he had done very poorly on it. With respect to the former objective, a number of techniques were used: In the first place, the scrambled sentence task was introduced as a test recently developed by the "Stanford people" in connection with a top level government project, a brief comment being added on the relationship of the Stanford people to the Stanford-Binet test. Second, *E* implied that the government project involved careful selection of personnel for critical intelligence work, and hence that the present test had been designed to provide precise discrimination in the average to superior range of adult intelligence. Third, bogus validity data were presented specifying academic and professional success as criteria. Finally, the title "Stanford Reasoning Test" was printed on the cover of the test booklet, and beneath that were printed "name," "age," and "IQ." (IQ was to be filled in, if known. Accomplices, discussed below, always filled in high IQs.) That an intelligence test would be administered in a study ostensibly concerned with "symbolic processes" was rationalized on the ground that E wished to compare the symbolic performances of average and highly intelligent students.

The following means were used to make *S* feel that he had done badly on the test: First, *E* stated that most Harvard students should perform very well on this test, since they were undoubtedly at the high end of the distribution of adult intelligence. Second, *Ss* were tested in groups of six, and each such group included two accomplices who pretended to solve all the tasks rapidly and exuded an air of confidence throughout the session. Finally, the four naive *Ss* received alternate forms of the test such that when any two of them failed to complete a task the other two *Ss* completed that same task, the result being that for any single task, four of the six testees achieved solution while two did not.

After completion of the exam, the second digit recall test was administered, and was immediately followed by a period of 8–10 minutes during the first part of which *E* made remarks intended to reactivate feelings of stress (referring to the exam's highly discriminative nature, the fact that scores would be made available, etc.), and during the remainder of which he scheduled *Ss* for subsequent sessions. Such scheduling was done at this point in order to make the interim period for the stress condition as long as that for the relief condition, and yet prevent *Ss* in the stress condition from discussing the exam among themselves. Half of the stress *Ss* were kept under stress for all three recall sessions, while the other half were relieved just prior to the second recall test.

Relief. This condition was exactly the same as the stress condition, except for the contents of the 8–10-minute interim period mentioned above. Rather than reactivate feelings of stress during this period, *E* attempted to relieve such feelings. He said the following: "Everybody sit back and relax. From the looks on your faces I gather that this was a pretty rough test—but perhaps you didn't do as badly as you think. You'll hate me when I tell you this, but believe me, it was all in the interest of science . . ." *E* then explained that the entire situation had been a hoax—that half the tasks had been unsolvable, that the two apparent geniuses in the group had really been *E*'s accomplices, and so on. This revelation released a flood of affect—best described as a mixture of surprise and tremendous relief—which was inevitably followed by a round-robin of descriptions of one's feelings and reactions during the session. The subsequent airing of these feelings gave the session a definite therapeutic tone.

Neutral. In contrast to the stress condition, the aim of the neutral condition was to create as informal and relaxed an atmosphere as possible, and to focus attention on the tasks rather than on *S*'s performance. To this end, the scrambled sentence tasks were introduced as a pretest of materials for a future experiment. The fact that the tasks were timed was rationalized by telling *S* that the purpose of the pretest was to eliminate tasks requiring too much time, *E* supposedly being interested in tasks that took no longer than about two minutes to solve. Moreover, whereas the timing for the stress condition was done in a test-like manner with *E* holding his stop watch conspicuously, the timing in the neutral condition was done unobtrusively and casually. The booklet of scrambled sentences, in addition, had no title on it, and *S* was told that he need not supply his name. Finally, no accomplices were present and all *Ss* had the same form of the test, so that everyone completed, or failed to complete, any given task. The 8–10-minute interim period between the second digit recall test and initial recall of the scrambled sentences was spent scheduling *Ss* for further sessions and having them fill out parts of some questionnaires.

RESULTS

Before considering the results which bear directly on our major questions, let us examine the data pertaining to a basic assumption of the study; namely, that the cognitive effects obtained in our stress condition were due to the ego-involving operations themselves, rather than to something inherent in the test materials or to subject differences alone. Comparison of the stress and neutral conditions for each of the subject groups indicates that this assumption is entirely tenable: Private and public school groups yielded significant differences ($p < .01$ in each case) between stress and neutral conditions on both the recall difference (Table 1, Column 2) and the per-

TABLE 1

MEAN INITIAL DIGIT RECALL (D), SCRAMBLED SENTENCE RECALL SCORES, AND PERCEPTUAL RECOGNITION SCORES FOR S-RECALLERS (PRIVATE SCHOOL *Ss*) AND F-RECALLERS (PUBLIC SCHOOL *Ss*) UNDER THREE EXPERIMENTAL CONDITIONS[a]

Condition	D	CR-IR	TR	CR	IR	t P^b
S-recallers (Private School *Ss*)						
Stress	9.5 (N = 19)	+1.1	5.5	3.3	2.2	+6.0
Relief	9.3 (N = 19)	+0.6	4.4	2.5	1.9	+0.9
Neutral	8.6 (N = 15)	−0.2	6.4	3.1	3.3	−1.7
F-recallers (Public School *Ss*)						
Stress	9.0 (N = 23)	+0.1	5.3	2.7	2.6	−2.6
Relief	10.2 (N = 23)	+0.1	5.7	2.9	2.8	−0.7
Neutral	9.4 (N = 18)	+1.2	6.8	4.0	2.8	+3.8

[a] The scrambled sentence recall and perceptual recognition measures include mean difference between number of completed and incompleted tasks recalled (CR-IR), mean total number of tasks recalled (TR), mean number of completed (CR) and incompleted (IR) tasks recalled, and mean perceptual threshold difference for threat-related and matched neutral words (P).

[b] A positive difference means higher thresholds for threat-related than for neutral words. The brightness difference in volts is twice the size of each entry.

ceptual threshold scores (Table 1, Column 6). Moreover, the directions across recall and perception were completely consistent for each group, private school *Ss* exhibiting S-recall as well as "repressive" perceptual patterns under stress but just the opposite tendencies under neutral conditions, public school *Ss* doing exactly the reverse. (The interaction Fs are significant beyond the .01 level for each variable. See Table 2.)[6,7]

For both private and public school *Ss*, total recall drops significantly from neutral to stress conditions (Table 1, Column 3). However, while for private school *Ss* this is due entirely to a decline in recall of failures (Column 5)[8] for public school *Ss* it is completely determined by a decline in recall of successes (Column 4). These losses assume added import when one compares the groups on our measure of recall ability—digit-retention. Column 1 of Table 1 shows the mean pretest digit-retention scores for the subgroups. The analysis of variance of these scores yielded no significant or near significant Fs (Table 2). Consider now the findings that indicate whether these losses are due to mnemonic or learning processes.

1. *Is the S-recall pattern a function of repression of failures or selective learning in favor of successes?* The critical datum here is the degree of shift in the recall difference score from stress to relief conditions for private school boys (Table 1, Column 2). Although a slight shift occurred, it was far from significant, t being less than 1. Recall of incompleted tasks (Table 1, Column 5) shows a nonsignificant decrease from stress to relief conditions, whereas, if a repression effect were operating and given the mnemonic comparability of our groups, one would expect, if anything, an increase (memory for failures being restored by the relief treatment). Recall of completed tasks (Column 4) and total recall (Column 3) also exhibit nonsignificant declines. In sum, then, no immediate recall shift appeared in the relief group. To determine, on the other hand, whether a delayed shift occurred for this group or for other groups; namely, for stress *Ss* who were subsequently relieved, and for nonrelieved stress *Ss* (the last as a function of time lapse alone), we turn to the data of Table 3 (Columns 1–6). None of the recall differences recorded there (for CR-IR, CR alone, or IR

[6] Lack of a strong recall tendency in favor of incompleted tasks in any of our subgroups indicates the presence of a general recall bias in favor of completed tasks. Thus, when the term "F-recall pattern" is employed in the present study, it is being used in a relative, not an absolute, sense. That this term is nevertheless being used appropriately, i.e., to specify a type of cognitive reaction, is indicated by the finding, noted earlier, that our group differences directly parallel those of Atkinson (3) for high and low need achievers tested under similar conditions.

[7] In all our analyses of variance, the form of the test, i.e., which particular tasks were completed and which were incompleted, was found to be a nonsignificant source, and hence its variance was added to the within-cells variance.

[8] These results are similar to those of Glixman (9). Whereas Glixman, however, regards such findings as indicative of a repression process, we are suggesting that they do not point unequivocally to such a conclusion.

TABLE 2

ANALYSES OF VARIANCE OF INITIAL DIGIT RECALL (D), SCRAMBLED SENTENCE RECALL SCORES, AND PERCEPTUAL RECOGNITION SCORES FOR S-RECALLERS (PRIVATE SCHOOL *Ss*) AND F-RECALLERS (PUBLIC SCHOOL *Ss*) UNDER THREE EXPERIMENTAL CONDITIONS

Source	D			CR-IR		TR		CR		IR		*P*	
	df	*ms*	*F*	*ms*	*F*	*ms*	*F*	*ms*	*F*	*ms*	*F*	*ms*	*F*
Type of *Ss*	1	3.86		0.76		5.80		0.56		2.62	1.19	111.34	2.15
Conditions	2	5.78	1.16	0.96		23.25	3.82*	7.46	3.50*	4.84	2.20	13.12	
Ss × Conditions	2	6.32	1.27	13.57	5.19*	5.50		5.38	2.53	4.55	2.07	458.50	8.84*
Error	110	4.99		2.61		6.09		2.13		2.20		51.87	

* For 2 and 110 *df*, an *F* of 3.09 is significant at the .05 level; an *F* of 4.81 is significant at the .01 level.

TABLE 3

MEAN DIFFERENCE BETWEEN FIRST AND SECOND RECALL SESSIONS, AND BETWEEN FIRST AND THIRD RECALL SESSIONS, FOR CR-IR, CR, AND IR

Condition of Initial Recall	Condition of Delayed Recall	S-recallers (Private School *Ss*)						F-recallers (Public School *Ss*)					
		2nd–1st sessions			3rd–1st sessions			2nd–1st sessions			3rd–1st sessions		
		CR-IR	CR	IR	CR-IR	CR	IR	CR-IR	CR	IR	CR-IR	CR	IR
stress	stress	−0.1 (N = 13)	0.2	0.3	−0.5 (N = 12)	0.0	0.5	−0.7 (N = 10)	0.2	0.9	0.1 (N = 10)	0.2	0.1
stress	relief	0.0 (N = 6)	0.0	0.0	−1.2 (N = 6)	−1.0	0.2	0.0 (N = 13)	0.3	0.3	0.4 (N = 13)	0.3	−0.1
relief	relief	0.1 (N = 19)	0.3	0.2	−0.2 (N = 16)	−0.2	0.0	−0.3 (N = 23)	0.0	0.3	−0.3 (N = 23)	−0.2	0.1
neutral	neutral	0.2 (N = 15)	0.4	0.2	0.2 (N = 15)	0.1	−0.1	−0.5 (N = 18)	0.0	0.5	−0.6 (N = 18)	−0.2	0.4

alone) are significantly different from zero,[9] hence implying no delayed restoration of forgotten items. The present findings thus do not support predictions derived from a repression interpretation of the S-recall pattern, but are consistent with those derived from a selective learning interpretation.

2. *Is the F-recall pattern a function of selective remembering or selective learning in favor of failures?* To answer this question, we must examine the degree of shift in the recall difference score from stress to relief conditions for public school *Ss* (Table 1, Column 2). No such shift occurred, the two means being identical. Recall of incompleted tasks alone, moreover, shows a nonsignificant increase, whereas the task-tension theory would predict, if anything, a decrease. Recall of completed tasks and total recall also show negligible increases. Here again, then, no immediate recall shift occurred in the relief condition. Further, there was no delayed shift for any of the subgroups. As indicated in the last six columns of Table 3, none of the differences for CR-IR, CR alone, or IR alone, were significantly different from zero. Our findings, therefore, do not confirm predictions derived from a task-tension interpretation of the F-recall pattern, but once more are consistent with those implied by a selective learning view.

3. *Is there a mnemonic reaction to the test as a whole?* The data relevant to this question are the threshold shifts from stress to relief conditions for test-related words (compared to matched neutral words). According to Table 1, Column 6, private school *Ss* in the relief group show a significant return ($p < .05$, two-tail test) to the level obtained under neutral conditions, whereas public school *Ss* do not shift significantly. That the perceptual

[9] As can be seen from Columns 5 and 6, the CR-IR difference of −1.2 ($p < .075$) is a function of decreased recall of completed tasks rather than increased recall of incompleted tasks.

defense pattern dissipated for private school *Ss* when they were relieved, suggests that a genuinely repressive reaction—a memory rather than a learning phenomenon—occurred for these *Ss* with regard to the threatening situation as a whole. Since the perceptual vigilance reaction of the public school *Ss*, on the other hand, did not dissipate in the relief condition, it would seem that this reaction is due to a learning rather than a memorial process.

DISCUSSION

Three propositions summarize the present results. 1. *Both the S- and F-recall tendencies in the present study are due to a selective learning rather than a selective remembering mechanism.* The pertinent datum here is the following: Among the boys from private schools, i.e., those exhibiting an S-recall pattern, the tendency to recall failures was no greater for relieved than for nonrelieved *Ss*; likewise, among the boys from public schools, i.e., those exhibiting an F-recall pattern, the tendency to recall failures was no less for relieved than for nonrelieved *Ss*. 2. *A repression mechanism was found for S-recallers, but with respect to the stress situation as a whole rather than for specifically failed items.* The evidence here: Nonrelieved private school boys manifested perceptual defense for words related in a general way to the examination situation, whereas those who were relieved showed no such perceptual defense. 3. *Instead of a comparable process of enhanced retention of the total stress situation, F-recallers showed enhanced registration for the total situation.* Thus, although public school boys under stress were perceptually vigilant for test-related words (relative to public school boys under neutral conditions), this effect did not dissipate in the relief condition.

Before we can accept these statements, two critical procedural issues require reexamination: namely, (*a*) was the stress situation sufficiently threatening; and (*b*) if so, was the relief treatment really therapeutic? Consider first the question of threat.

Needless to say, the present data would have no bearing on a theory of repression if the stress situation had not induced intense ego threat. That such threat did indeed occur may be inferred from several lines of evidence—observational, introspective, and experimental. In the first place, *Ss* in the stress condition showed many overt signs of emotional disturbance (little or none of which appeared among the neutral *Ss*). When stumped by a problem, they would squirm nervously or shake their heads in disgust (sometimes swearing simultaneously) or would stare at the page with a very pained expression. Further, there were many attempts to conceal one's difficulty from others, either by pretending to have finished problems not actually completed, or, when able to solve a problem, by trying to make one's success very obvious to competitors (e.g., putting one's pencil down with a bang or yawning loudly while stretching). Again, depressive and hostile reactions were quite prevalent following the exam, many *Ss* sitting morosely in their chairs, others openly challenging the worth of the test. Lastly, expressions of great relief—deep sighs, postural relaxation, laughter—almost always appeared when the ruse was finally explained. As to introspective evidence: First, most *Ss* voluntarily reported that they had been extremely concerned prior to, and tremendously relieved following, exposure of the hoax. They attributed emotions to themselves ranging all the way from intense feelings of inadequacy ("I said to myself, 'See, you really don't belong at Harvard' ") to violent hostility ("I wanted to get that supercilious bastard sitting beside me.") Secondly, continued questioning revealed that although some thought the test a poor one, no one had doubted its reality. With regard to experimental data, finally, consider the results for our indicator of anxiety: digit recall. Whereas recall of digits showed a marked pretest to posttest improvement for *Ss* in the neutral condition, *Ss* under stress showed no such improvement, the *F* between conditions being significant well beyond the .01 level. The most plausible explanation for the latter *Ss*' deficiency would seem to be the disrupting effects of anxiety generated during the scrambled sentence test. These data, in sum, would seem to support the assumption that intense ego threat was aroused in the present study.

Let us turn now to the second question: Can exposure of the hoax really be considered a therapeutic, i.e., anxiety-reducing, technique?

Even if it were conceded that anxiety had been aroused in our stress condition, the present findings could be legitimately challenged on the ground that this anxiety was not effectively reduced in the relief condition. Such a challenge might proceed as follows: In the typical clinical setting, the therapist's efforts to uncover repressed events meet with strong defensive resistance. Hence, recovery of repressed material requires a great deal of time and involves a special kind of intimacy between patient and therapist, neither of which occurred in the present relief treatment. The point is well taken, but it overlooks two important differences between repression as it ordinarily occurs and repression as precipitated in the present experiment. First, in ordinary circumstances, the fearful event is unknown to the therapist, having occurred before therapist and patient began their relationship. Thus, as long as the therapist is ignorant of the actual situation and cannot confront the patient with reality, the patient may safely continue to hide behind his defense and deny the event. In the present situation, however, the individual is faced with an omniscient therapist—one who produced the threatening event and who is as much an authority on its nature as the patient. To continue repressing in the face of such omniscience is to render one's contact with reality precarious at best. Second, and perhaps more critical, there would seem to be but two ways in which a threatening event may be made less fearful: (*a*) by effecting a radical change in the patient's motivational structure (e.g., eliminating his fear of intellectual incompetence) so that the event (e.g., exam failure) no longer poses a threat; or (*b*) by altering the significance of the event so that it is no longer motivationally relevant. The former is the proper goal of the therapist in the usual clinical setting. The latter is what was obviously involved in the revelation of the experimental hoax to our *S*s: the failure experience had initially been created by deception and was now being restored to its true meaning as a psychological gimmick. *E*'s revelation does not make *S* any less of a repressor or eliminate his problem with respect to intellectual competence (nor was it intended to), but merely renders this particular event irrelevant to that problem and hence no longer something to be defended against. In the light of both these points, then, the assumption that effective therapy occurred in the present experiment seems entirely tenable.

Two general conclusions follow from the present findings. One is methodological, and touches on the question of criteria for a suitable test of repression. The other concerns the particular mechanisms that underlie the avoidant and adient cognitive reactions obtained here. Consider each of these in turn.

1. In attempting to provide an experimental demonstration of repression, the general practice has been to compare the recall of certain "nonthreatening" aspects of the experimental situation with that for certain threatening aspects. In particular, it has been common to examine the recall of failures in contrast to successes on a bogus intelligence test. The present results impute this approach, for they suggest that repression embraces only the most general aspects of the stress situation, whereas recall of particular threatening or "nonthreatening" aspects of that situation is a function of learning. This finding seems entirely reasonable, for when a person has failed half the items on a test, what is likely to be threatening for him is not that he has failed those particular items but that he has failed the *test*. Nor can the items he has passed be entirely free from threat if the total situation in which they are embedded arouses anxiety. It would thus seem necessary, in experimental research on repression, to determine precisely what category of materials is relevant to the subject's conflicts and fears. While recent experiments in this area have been careful to satisfy the previously neglected criterion of motivational relevance of experimental materials, the present findings strongly suggest it would also be well to consider the cognitive relevance of these materials.

2. That both public and private school *S*s showed opposite and significant differences between stress and neutral conditions, and that the directions of these differences were completely consistent across recall and perception, indicate that two quite different defensive reactions to stress occurred in this study. One—an avoidant reaction by private school *S*s—was indeed shown to be repressive as far as the total test situation was concerned, but, with regard to particular items, was a

matter of registration differences. The other—an adient reaction by public school *Ss*—was a case of registration differences at both levels. One may properly inquire at this point what the particular processes might be which determine the various registration differences found. Osgood (15), among others, has suggested that differential rehearsal may account for such learning superiorities. The possibility of covert review of failures by Harvard public school *Ss* indeed seems especially reasonable on the assumption that these *Ss* are high achievers. While such individuals are not unconcerned about failure, they undoubtedly have a history of repeated reward for efforts to counteract initial failure. Hence these *Ss* would be expected to ruminate over their difficulties as a cognitive precursor to constructive adjustment. In the case of private school *Ss*, covert review of successes might be expected as a kind of ego-bolstering reaction. Another possibility for these *Ss* is that anxiety over inability to solve items acts to disrupt their registration. These two mechanisms are, of course, by no means mutually exclusive.

SUMMARY

The present study proposed to answer three questions: 1. Is superior recall of successes in an intelligence test situation (S-recall) a function of selective forgetting (repression) of failures or selective learning in favor of successes? 2. Is superior recall of failures in this situation (F-recall) a function of selective remembering or selective learning in favor of failures? 3. Is there a mnemonic reaction to the test as a whole? The first two questions were examined by comparing the recall preference of nonrelieved S- and F-recallers on a bogus intelligence test (involving solvable and unsolvable scrambled sentences) with that for comparable relieved *Ss*, selective remembering theories predicting a shift from nonrelieved to relieved conditions for each recall group. The third question was examined by comparing the perceptual recognition thresholds of the above relieved and nonrelieved *Ss* on exam-related and matched neutral words. Forty-two *Ss* each were tested in the relief and nonrelief condition, with 33 more tested under neutral conditions. Prior attendance at a private versus a public high school was used as the indicator of S- and F-recallers respectively.

The findings were as follows: 1. Both the S- and F-recall tendencies were due to a selective learning rather than a selective remembering mechanism. 2. A repression mechanism was demonstrated for S-recallers with regard to the stress situation as a whole. 3. A comparable process of enhanced retention of the total stress situation was not demonstrated for F-recallers; rather, the effect here was due to enhanced registration. Two conclusions were drawn from these findings: First, that repression was found for the total test situation rather than for specific items, suggests experimental tests of repression must fulfill a criterion of cognitive relevance of test materials. Second, the kinds of enhanced registration found may have been due to covert rehearsal, and possibly also to disruption of the registration of other items by anxiety.

REFERENCES

1. ALPER, THELMA G. Memory for completed and incompleted tasks as a function of personality: An analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41,** 403–420.
2. ALPER, THELMA G. Memory for completed and incompleted tasks as a function of personality: correlation between experimental and personality data. *J. Pers.*, 1948, **17,** 104–137.
3. ATKINSON, J. S. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46,** 381–390.
4. BLUM, G. S. *Psychoanalytic theories of personality.* New York: McGraw-Hill, 1953.
5. CARON, A. J. The effects of task, objective orientation, and personality factors on recall in an interrupted-task situation. Unpublished doctoral dissertation, Univer. of Chicago, 1954.
6. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47,** 230–235.
7. ERIKSEN, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952, **44,** 442–447.
8. ERIKSEN, C. W. Psychological defenses and "ego strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49,** 45–50.
9. GLIXMAN, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39,** 281–295.
10. GUILFORD, J. P. *An Inventory of Factors STDCR.* Beverly Hills, Cal.: Sheridan Supply, 1940.
11. KNUTSON, A. L. Personal security as related to station in life. *Psychol. Monogr.*, 1952, **66,** No. 336.
12. LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

13. MCARTHUR, C. Personalities of public and private school boys. *Harvard Educ. Rev.*, 1954, **24**, 256–262.
14. MCARTHUR, C. Personality differences between middle and upper classes. *J. abnorm. soc. Psychol.*, 1955, **50**, 247–254.
15. OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.
16. RAPAPORT, D. *et al. Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.
17. ROSENZWEIG, S. An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64–74.
18. ROSENZWEIG, S., & SARASON, S. An experimental study of the triadic hypothesis: reaction to frustration, ego defense, and hypnotizability. I. Correlational approach. *Charact. & Pers.*, 1942, **11**, 1–20.
19. SCODEL, A. Passivity in a class of peptic ulcer patients. *Psychol. Monogr.*, 1953, **67**, 1–15.
20. UNDERWOOD, B. J. Learning and retention. In C. P. Stone (Ed.), *Annual review of psychology*, Vol. III. Stanford: Annual Reviews, 1953.
21. WECHSLER, D. *The measurement of adult intelligence*. (3rd ed.) Baltimore: Williams & Wilkins, 1944.
22. ZEIGARNIK, BLUMA. Ueber das Behalten von erledigten und unerledigten Handlungen. *Psychol. Forsch.*, 1927, **9**, 1–85.
23. ZELLER, A. F. An experimental analogue of repression. I. Historical summary. *Psychol. Bull.*, 1950, **47**, 39–51.

Received June 17, 1957. Prior Publication.

GROUP CLIMATE AS A FACTOR IN THE RELEVANCE AND ACCURACY OF SOCIAL PERCEPTION[1]

RALPH V. EXLINE

Fels Group Dynamics Center, University of Delaware

THIS PAPER reports on a study of accuracy of social perception as a function of the "climate" of social interaction. Special attention is paid to the relevance of the perceived, estimated, or judged social behavior of others to the goals of the perceiver.

Much of the research on accuracy of social perception has been concerned with accuracy as an attribute of the individual perceiver (3, 8, 12). Results of certain of these studies, however, suggest the merit of studying the relationship between accuracy of perception and the social context in which perception is measured. Conrad (5) and Chowdhry and Newcomb (4), for example, suggest that accuracy is a function of the relevance of the perceptual task to the motives of the perceiver, while Suchman (22) speculated that group differences in accuracy of task-oriented perception are caused by the nature of the interaction within the groups studied.

The theoretical relationship often assumed to exist between motivation and perception suggests that the study of perceptual accuracy in a social context may be facilitated by considering perception in terms of a path-goal relationship. Thus, according to Lewinian theory (16), once a person defines a goal, he attempts to move from a position outside to a position within the goal region. In a relatively unstructured situation where a goal, but not the means to achieve it, has been defined, movement toward the goal requires that he direct energy to the differentiation of a path leading toward the goal. This thinking can be generalized to group situations and seems especially pertinent where information concerning the opinion of other group members is a necessary part of the path region (and where the opinion of others can be learned, or inferred, from their behavior). In these cases, persons can be expected to direct energy to the understanding of relevant opinions. Hence, everything else being equal, they should be more likely to be aware of and to perceive accurately those behaviors of others which they define as relevant to locomotion toward the goal.

In the light of research and theory (7, 16, 18) that conceives of the group as providing a psychological social environment for its members, the environment created by the group may have predictable effects on the likelihood that a given behavior is seen as a "path" element and hence on the accuracy with which it is perceived. It is the general hypothesis of this study that the group environment, through its effect upon what is perceived as relevant to the goals of the group members, predictably affects the accuracy with which specified social phenomena are perceived.

One aspect of the psychological social environment that groups provide their members is that of the social "climate" (18) or "hedonic tone" (15) that characterizes the interaction of group members. The relationship of this phenomenon to goal-striving activity of individuals has been widely documented. Research reported by Lippitt and White (18), Deutsch (7), and Hemphill and Westie (15) has shown that friendliness and congeniality of the group's "social climate" occurs concomitantly with motivation to persist in working toward task goals assigned to the group. There is also evidence to indicate that individuals as members of groups often pursue individual, or egocentered goals, rather than ones oriented to the group task. The social climate of these latter groups has been shown to be one of less pleasantness (13), more conflict (10), and less friendliness (7) than the climate of groups where members' goal-orientations are less individual in nature.

The effect of the group's climate upon the relevance of a given goal, hence upon the nature of behavior seen as defining the goal path, provides the bridge between the group environment and the accuracy with which individuals perceive specified behaviors. Thus, if members of congenial groups are more likely

[1] This report is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Illinois.

to hold a group goal in common, opinions that bear on the group goal are more likely to be mutually relevant, hence more accurately perceived, than in groups of less congenial atmosphere. Conversely, if members of groups characterized by uncongenial interaction climates are more prone to hold individual, ego-centered goals, then hypothetically, opinions held and communicated by others that are relevant to one's own ego-centered goals should be more relevant, hence more accurately perceived.

On the assumption that task-oriented activity is perceived to be more relevant in congenial than in noncongenial group climates, while ego-oriented activity is perceived to be more relevant in noncongenial than in congenial group climates, the discussion above may be more formally stated in terms of two hypotheses which this study is designed to test.

Hypothesis I: *Members of congenial groups are more accurate than members of noncongenial groups in perceiving task-oriented behaviors.*

Hypothesis II: *Members of noncongenial groups are more accurate than members of congenial groups in perceiving ego-oriented behavior.*

METHOD

Subjects

One hundred thirty graduate students assumed to be relatively homogeneous as to age, intelligence, and socioeconomic status (they were public school teachers) were recruited from a summer session at the University of Illinois. To minimize the effects of previous acquaintance, each group was composed of *S*s drawn from a different section of a basic course in educational psychology. The fact that these students were teachers whose regular jobs were in communities scattered throughout the state was taken to be an additional restriction on acquaintance.

Design

The *S*s were randomly assigned to two treatments by group atmosphere, then subdivided into groups of five people of the same sex. The groups were first given instructions designed to create a congenial or noncongenial atmosphere. Following these instructions, each group worked on a standard discussion task, after which the members marked a set of rating scales designed to measure the accuracy of their perception of task-oriented and ego-oriented behaviors of others, their satisfaction with the progress of the group, and their willingness to continue working with the group.

Operational Definitions[2]

Congenial groups.[3] Congenial groups were those groups which, by a process of random selection, were given information designed to create the expectation that group members would be congenial to one another. At the time of recruitment, *S*s were given a standard personality test and a questionnaire purported to measure attitudes towards working in groups, and told that this information would be used to help the investigator assign them to discussion groups in which they would find the other members congenial. Upon arriving at the laboratory, *S*s assigned to the congenial treatment were told that they stood very high on a scale of congeniality developed from the preliminary tests—"the matching was so good as to be expected only once or twice in all the groups studied," and that they should get on well together.

Noncongenial groups. These groups were given information designed to create the expectation that group members would not be congenial to one another. The *S*s assigned to these groups were told that due to the limited number of available subjects, not all groups could be matched equally well as to congeniality. They were told that they stood very low on the congeniality scale, that it was necessary to put people with opposing views in the same group, that they were about as bad a matching as was possible, and that they probably would not find their group very congenial.

Task. The task was introduced as one that would enable psychologists to relate the *S*s' scores on the previous tests to their ability to take the role of another person. It was implied that psychologists think this ability important in establishing good human relationships with other people.

The *S*s were asked to role-play a group of trainees in vocational guidance, meeting to discuss the disposition of a vocational counselee. They were to reach agreement upon a recommendation concerning the disposition of the case. Each *S* was given the identical test and biographical information about the hypothetical client, but in addition received a unique recommendation based on one of five different orientations toward the data. Thus, one *S* suggested a job on the basis of an orientation that emphasized the need to consider present and future job opportunities. Another *S*'s suggestion was based on the relationship between personality factors and job requirements, a third on the role of training and past experience, a fourth on test scores, and the fifth on test scores in combination with training and past experience. The disagreement that was thus ensured was expected to facilitate the production of behavior indicating interpersonal preference as well as of differential orientation toward the path to task completion.

A standard amount of time was allowed for *S*s to

[2] Complete descriptions of the instructions, results by individual *S*s and detailed statistical tables can be obtained in reference (9) from University Microfilms, Ann Arbor, Michigan.

[3] Instructions designed to create congenial and noncongenial groups were a modified version of those used by Back (1).

familiarize themselves with the data and the point of view they were to support in the discussion. To insure that each group had a minimum opportunity to hear all points of view, the instructions required each person to present his suggestion and opinion in turn. After each *S* had spoken twice, a general discussion was permitted.

Task-oriented behavior. Task-oriented behavior is defined conceptually as the manipulation, symbolic or otherwise, of the idiosyncratic content of the task. Its operational definition, like that of ego-oriented behavior, becomes complex when the individual is a member of a problem-solving group, for the definition must involve reaction to others' acts. Action and reaction are necessary to the processes of communication, evaluation, control, decision, and tension reduction which according to Bales (2) characterize the sequence of group problem solving. Individuals in such groups communicate not only their own orientations to the problem, but their evaluations of other's orientations. Both such types of communications qualify as task-oriented behaviors in that both are necessary to the movement of the group toward its goal of decision. On the basis of these considerations, task-oriented behavior is operationally defined for present purposes as verbal or nonverbal behavior that communicates either the frame of reference used by a person to organize the task data, or his reaction to the frames of reference used by others to organize the same data.

Ego-oriented behavior. The individual is said to be ego oriented when his goal is the enhancement of self without regard to the demands of the task environment (17). Ego-oriented behavior refers to the manipulation of elements in the task situation that bear on the self as a value. It is assumed that in the process of working through disagreements, *S*s communicate information pertinent to their orientation to and evaluation of their fellow group members as persons. Ego-oriented behavior, then, is operationally defined as verbal or nonverbal behavior that communicates a person's relative acceptance of other group members.

Accuracy of perception of task-oriented behavior. Task-oriented perceptual accuracy is defined as the discrepancy between each *S*'s rank ordering of the use of four orientations toward counseling as he thinks the other *S*s would rank order them, and the other *S*'s actual rank ordering of the same four orientations.

An error score for each *S* was obtained as follows. Each *S* was given a booklet containing the definitions of the orientations toward counseling and five sets of rating scales. Each scale consisted of 20 intervals distributed over six definition points from "never used" to "exclusive use." Each *S* filled out one set of ratings to indicate the extent to which he felt he had used the four independent orientations[4] in his own discussion. Each *S* also filled out the remaining sets of scales to indicate his perception of the rank order of the extent to which each of the others favored the various orientations. Thus, each *S*'s own responses provided a key against which the estimates made by the others could be scored for errors. An over-all error score representing *S*'s ability to accurately predict the rank order assigned to each orientation by each *S* was obtained by averaging *S*'s squared errors over all orientations and others.

Accuracy scores defined in terms of a rank-order metric have the advantage of eliminating the possibility that low error scores merely reflect the judges' willingness to use extremes, or ability to judge deviations of others from a central tendency. Scores derived by this method reflect a component of accuracy that Cronbach (6) terms SAr, i.e., accuracy in judging the rank ordering of the items made by the group as a "generalized other." Since items correspond to task orientations, this type of accuracy is assumed to be relevant to the problem studied.

Accuracy of perception of ego-oriented behavior. Ego-oriented accuracy was measured in a manner similar to that of task-oriented accuracy. In this case, the *S* rated the extent to which he liked the others and estimated the ratings each of the others gave their co-workers. Ratings were made on scales which, but for re-worded designation points, were identical with those used for task-orientation. An error score was derived as described in task-oriented perception. In this case, however, the other members of the group were the items which *S*s rank-ordered.

RESULTS

Effectiveness of Manipulations

To establish that the treatments did indeed create the necessary differences in congeniality, the extent to which members of the various groups liked their co-workers, and their willingness to continue with the group, were assessed. Members of "congenial" groups marked a scale of liking of co-workers significantly higher than did the members of "noncongenial" groups (means of 16.05, 14.60; *SD*s, 1.99, 1.95; $p = .02$, two tails). Similarly, the "congenials" rated themselves significantly higher ($p = .02$) in willingness to continue working with the group (means of 16.72, 15.19; *SD*s, 3.26, 3.92). Since *S*s assigned to congenial groups showed no greater pre-experimental disposition to work with groups than did *S*s assigned to noncongenial groups[5], the above data are taken as evidence that the inductions created the desired climates of congeniality.

[4] The orientation of the fifth *S* was a combination of those of the third and fourth *S*s.

[5] The conclusion that *S*s in congenial and noncongenial groups did not differ in their pre-experimental disposition to work in groups, was based on the response to an item in the questionnaire, given to *S*s in their classes, which measured attitudes toward working in groups. A low score on this item signified greater willingness to work in groups. Means of *S*s in congenial and noncongenial groups were, respectively, 2.41 ($N = 65$, $SD = .94$) and 2.35 ($N = 59$, $SD = .92$). Six *S*s in noncongenial groups did not answer the question. A t of .97 significant at greater than the .30 level (two-tailed test) was obtained.

TABLE 1
MEAN ERROR SCORE IN ESTIMATING TASK- AND EGO-ORIENTED BEHAVIOR MADE BY SUBJECTS CATEGORIZED AS TO CONGENIALITY TREATMENTS AND SEX

Perception Measure	Treatment			
	Congenial		Noncongenial	
	N	$\bar{X}$	*N*	$\bar{X}$
Task-oriented				
Male subjects	32	7.16	32	8.06
Female subjects	32	6.51	32	7.89
Ego-oriented[a]				
Male subjects	20	9.55	25	10.11
Female subjects	30	7.98	19[b]	8.96

[a] Mean-per-person error is based on five-person groups only. The difficulty of this estimation was not comparable in the five- and four-person groups.
[b] One *S* filled out scales improperly.

Tests of Hypotheses

The hypotheses concerning the relationship between social climate and the accuracy of perception of task and ego-oriented behavior of others were tested by applying analysis of variance techniques to mean-per-person error scores. Although the problem was not formulated in terms of sex differences, the composition of the groups permits this comparison.

Hypothesis I. Mean-per-person error scores on task-oriented perception for congenial and noncongenial groups are listed in Table 1 and the results of the analysis of variance in Table 2. The results indicate that members of congenial groups, regardless of sex, are significantly more accurate than members of noncongenial groups in task-oriented perception.

Hypothesis II. Mean-per-person error scores on ego-oriented perception for congenial and noncongenial groups are listed in Table 1 and the results of the analysis of variance in Table 3. An examination of ego-oriented perception error scores shows no statistically significant differences between congenial and noncongenial groups, although significant differences attributable to sex are readily apparent.

Hypothesis I is thus confirmed by the data, while Hypothesis II is not. Indeed, the data indicated a trend, albeit a statistically insignificant one, in the direction of greater ego-oriented accuracy on the part of members of congenial groups. An unanticipated but interesting finding is that women are more accurate than men in the estimation of their co-workers' interpersonal likes and dislikes.

TABLE 2
ANALYSIS OF VARIANCE OF TASK-ORIENTED ESTIMATION ERROR OF SUBJECTS CATEGORIZED BY SEX AND CONGENIALITY TREATMENT

Source of Variation	*df*	Mean Square	*F*
Sex	1	5.36	.85
Congeniality treatment	1	41.63	6.62*
Interaction: sex × congeniality	1	1.95	.31
Within groups	124	6.29	

* Significant at .05 level or beyond.

TABLE 3
ANALYSIS OF VARIANCE OF EGO-ORIENTED ESTIMATION ERROR OF SUBJECTS CATEGORIZED BY SEX AND CONGENIALITY TREATMENT

Source of Variation	*df*[a]	Mean Square	*F*
Sex	1	43.47	4.90*
Congeniality treatment	1	14.33	1.62
Interaction: sex × treatment	1	.71	.08
Within groups	90	8.87	

* Significant at the .05 level or beyond.
[a] Tsao's (24) approximation method was applied to correct for disproportionate subclass n.

Post Hoc Analyses

Examination of *S*s' responses to the post-meeting questionnaire provide additional data concerning the relationship between group climate and accuracy in perceiving task and ego-oriented behavior. It has already been mentioned that *S*s were asked to indicate their willingness to continue working with the group on similar tasks in order to measure the effect of the congeniality inductions. If this measure is a valid index of congeniality, *S*s more willing to continue should perceive task-oriented behavior more accurately than do those who find the group less attractive, regardless of experimental treatment; while, conversely, those less willing to continue should perceive ego-oriented material more accurately.

These *post hoc* hypotheses were tested by first categorizing *S*s according to whether their willingness to work with the same group was above or below the median of the pooled scores, then applying analysis of variance techniques to determine the effects of congeniality inductions, willingness, and the interaction between the two upon the social perception error scores. Mean-per-person error scores are listed in

TABLE 4

MEAN ERROR SCORE IN ESTIMATING TASK- AND EGO-ORIENTED BEHAVIOR MADE BY SUBJECTS CATEGORIZED AS TO WILLINGNESS TO CONTINUE AND CONGENIALITY

Perception Measure	Treatment			
	Congenial		Noncongenial	
	N	$\bar{X}$	*N*	$\bar{X}$
Task-oriented				
High willingness	37	6.67	27	7.13
Low willingness	27	7.06	37	8.58
Ego-oriented				
High willingness	29	9.45	18	10.04
Low willingness	21	7.89	26	8.97

TABLE 5

ANALYSIS OF VARIANCE OF MEAN ERROR SCORES IN TASK-ORIENTED PERCEPTION MADE BY MORE AND LESS WILLING SUBJECTS

Source of Variation	*df*[a]	Mean Square	*F*
Willingness	1	27.08	4.593*
Congeniality treatment	1	31.36	5.318*
Interaction: willingness × congeniality	1	8.99	1.524
Within groups	124	6.12	

* Significant at the .05 level or beyond.

[a] Tsao's (24) approximation method was used to correct disproportionate subclass n.

Table 4 and the results of the analysis of variance in Tables 5 and 6.

Inspection of the means in Table 4 shows that, in both congenial and noncongenial groups, *Ss* who were more willing to continue with the group made fewer task-oriented errors than did their less willing co-workers. It is also apparent that in both congenial and noncongenial groups, fewer errors in assessing interpersonal relations were made by those less willing to continue. Since the *F* ratios in Tables 5 and 6 indicate that effects of willingness are significant at better than the .05 level, the data support the hypotheses.[6]

[6] Although no data concerning the productivity of the groups were collected, it is of interest to note that members of congenial groups expressed greater satisfaction with the progress of their groups. Mean satisfaction ratings, on a 20-interval scale similar to those described previously, were noted as follows: Congenial male and female, 14.44 and 15.47 respectively; noncongenial male and female, 12.19 and 11.06 respectively. With mean squares of 354.44, .07, 37.19, and 18.17 for congeniality, sex, interaction, and error variance, respectively, the only significant *F* was that of 19.51 (significant at the .001 level with *df*'s of 1 and 124) found for congeniality. It is uncertain whether the difference noted above is related to the greater task-oriented accuracy recorded in congenial groups or merely reflects the impact of the congeniality inductions. In view of this uncertainty these scores are not discussed further.

TABLE 6

ANALYSIS OF VARIANCE OF MEAN ERROR SCORE IN EGO-ORIENTED PERCEPTION MADE BY MORE AND LESS WILLING SUBJECTS

Source of Variation	*df*[a]	Mean Square	*F*
Willingness	1	40.64	4.089*
Congeniality treatment	1	16.38	1.684
Interaction: willingness × congeniality	1	1.41	.014
Within groups	90	9.94	

* Significant at the .05 level or beyond.

[a] Tsao's (24) approximation method was used to correct disproportionate subclass n.

DISCUSSION

Results reported above call into question the generality of a theory of accuracy of social perception that postulates accuracy as a function of the relevance of the perceived behavior to the goals of the perceivers. On the one hand, the finding that task-oriented phenomena are more accurately perceived in congenial group atmospheres, where they are assumed to be more relevant to the goals of group members, would appear to provide support for the theory. On the other hand, the theory was not supported by the finding that ego-oriented phenomena are not more accurately perceived in the noncongenial group condition, the setting where ego-orientation was assumed to be more relevant.

It is possible that the operational measure of ego-oriented accuracy of perception did not provide a valid test of the theory with reference to such behavior. There is evidence to suggest (1, 2, 13) that the mapping of interpersonal relations is a relevant activity for members of congenial groups.

On the other hand, existing theory and data suggest that regardless of the validity of operationally defining ego-relevant activity as the assessment of interpersonal relations, one could not expect such assessment to be accurately carried out. Newcomb's (19) theory of communicative acts, for example, suggests that accuracy of perception of others' behavior in a group is a function of the extent to which

the behavior can be said to be an object common to group members. Thus, while it is possible to conceive of task-oriented behavior as an object common to the members of congenial groups, it is less easy to see how others' behavior perceived as relevant to the private, ego-oriented goals of one group member can be said to be held in common by all group members. Also relevant to this discussion are data reported by Tagiuri, Blake, and Bruner (23), who indicated *S*s could not accurately perceive cues of rejection. To the extent that such cues are more likely to occur in noncongenial than in congenial groups, a factor favoring greater accuracy of perception of interpersonal relations exists in congenial groups.

In spite of the above arguments, there remains the fact that interpersonal relations were perceived with greater accuracy by those less willing to continue their association with the group. It was suggested, moreover, that this result could be interpreted as supporting the hypothesized relationship between the congeniality of the group and accuracy of perception of ego-relevant behavior.

An alternate interpretation may explain this relationship. This interpretation, which assumes that interpersonal conflict occurs in the groups, is at variance with the specific derivation embodied in Hypothesis II, but not necessarily with the general theoretical position of this paper (conflict, after all, is a feature of the group's environment). If cues concerning interpersonal relations are more likely to be observed in situations of interpersonal conflict, and if one can assume that conflict adversely affects a member's willingness to continue with the group, it follows that accuracy in perception of interpersonal relations and unwillingness to continue with the group may be concomitant results of conflict. Unfortunately, there are no measures of the extent to which such conflict marked the process of group discussion. The question of the meaning of more accurate perception of interpersonal relations on the part of those less willing to continue with the group thus remains unanswered. In view of the speculative, even *post hoc*, nature of the postulated relationships among willingness, congeniality, and accuracy of perception, it is suggested that generalization of the theory herein presented be limited to the perception of task-oriented phenomena. The significant relationships noted between willingness and the two kinds of accuracy of perception, however, do suggest the desirability of further exploring the nature of these relationships, as well as their implications for the processes of problem-solving groups.

Finally, the unanticipated finding that women perceive interpersonal relations more accurately than men appears to merit some discussion. This difference is not related to sex differences in willingness to work with the group, a factor already shown to be inversely related to such accuracy, for no differences between the willingness of men and women were found in either treatment. We must look further for an explanation of this phenomenon.

One explanation of this finding rests on the assumption that our culture sanctions freer expression of feeling by women than by men. It would follow, from this assumption, that more women than men would openly express their likes and dislikes concerning their coworkers. This explanation would appear to be related to the theory of sex-role differentiation developed by Parsons and Bales (20), and studies in jury deliberations by Strodtbeck and Mann (21). The results herein reported might be interpreted as further and more general confirmations of the above theory concerning woman's sex role.

An alternative explanation might be phrased in terms of social power. If the general position of women in our society is one which allows them less social power than men, it may be important for them to be able to accurately assess the feelings, attitudes, and relationships among others in order to achieve their personal goals.

Finally, women may be higher in "affiliation motivation" than men. If this is so, French and Chadwick's (11) finding that those having high affiliation motivation are more accurate in perceiving their own popularity status would be pertinent to our data. Research designed to test the validity of the above speculations is in progress.

SUMMARY

Twenty-eight groups of five or four graduate students each, controlled for sex, education, and acquaintance were assigned to congenial or noncongenial group climates. Analysis of variance was used to test the effects of group

climates, motivational levels, and the importance of sex difference upon perceptions of task and ego-oriented behavior.

The findings were as follows:

1. Individuals in congenial groups were significantly more accurate in perceiving task-oriented behavior of their group than were members of noncongenial groups.

2. Individuals in noncongenial groups were not more accurate in perceiving the interpersonal relations in the groups than were individuals in congenial groups.

3. Individuals willing to continue to work with the group on a similar task were more accurate in perceiving task-oriented behavior of their group than were individuals less willing to continue with the group.

4. Individuals willing to continue to work with the group on a similar task were significantly less accurate in perceiving interpersonal relations within their group than were individuals less willing to continue with the group.

5. Women, though neither more nor less willing to continue with their group, were significantly more accurate than men in perceiving interpersonal relations.

It is concluded that social climates can predictably affect the perception of task-oriented behavior variables, but that further research on the relationships among social climate, ego-oriented behavior, and accuracy in social perception is necessary. Research into the determinants of sex differences in perception of interpersonal relations is recommended.

REFERENCES

1. BACK, K. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.
2. BALES, ROBERT F. *Interaction process analysis.* Addison-Wesley Press, 1950.
3. BIERI, J. Cognitive complexity simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 263–268.
4. CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.
5. CONRAD, H. S. The personal equation in ratings: IIA systematic evaluation. *J. educ. Psychol.*, 1933, **24**, 39–46.
6. CRONBACH, L. J. Processes affecting scores on "Understanding of others" and "Assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–194.
7. DEUTSCH, M. The effects of cooperation and competition upon group process. In D. Cartwright and A. Zander (Eds.), *Group dynamics: Research and theory.* Evanston, Illinois: Row Peterson, 1953. Pp. 319–353.
8. DYMOND, ROSALIND F. A preliminary investigation of the relation of insight and empathy. *J. consult. Psychol.*, 1948, **12**, 228–233.
9. EXLINE, R. V. Relevance as a determiner of the difficulty and validity of social perception. Unpublished doctoral dissertation, Univer. of Illinois, 1954.
10. FOURIEZOS, N. T., HUTT, M. L., & GUETZKOW, H. Self-oriented needs in discussion groups. *J. abnorm. soc. Psychol.*, 1950, **45**, 682–690.
11. FRENCH, E. G., & CHADWICK, I. Some characteristics of affiliation motivation. *J. abnorm. soc. Psychol.*, 1956, **52**, 296–300.
12. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
13. GERARD, H. B. The anchorage of opinions in face-to-face groups. *Hum. Relat.*, 1954, **7**, 313–326.
14. GUETZKOW, H., & GYR, J. An analysis of conflict in decision-making groups. *Hum. Relat.*, 1954, **7**, 367–382.
15. HEMPHILL, J. K., & WESTIE, C. The measurement of group dimensions. *J. Psychol.*, 1950, **29**, 325–342.
16. LEWIN, K. Behavior and development as a function of the total situation. In D. Cartwright (Ed.), *Field theory in social science.* New York: Harper, 1951.
17. LEWIS, HELEN B. An experimental study of the role of the ego in work. *J. exp. Psychol.*, 1944, **34**, 113–126.
18. LIPPITT, R., & WHITE, R. The "social climate" of children's groups. In R. Barker, J. Kounin, & H. F. Wright (Eds.), *Child behavior and development.* New York: McGraw-Hill, 1943.
19. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.
20. PARSONS, T., & BALES, R. F. *Family socialization and interaction process.* Glencoe, Ill.: Free Press, 1955.
21. STRODTBECK, F. L., & MANN, R. D. Sex role differentiation in jury deliberations. *Sociometry*, 1956, **19**, 3–11.
22. SUCHMAN, J. R. Social sensitivity in the small task-oriented group. *J. abnorm. soc. Psychol.*, 1956, **52**, 75–83.
23. TAGIURI, R., BLAKE, R., & BRUNER, J. S. Some determinants of the perception of positive and negative feelings in others. *J. abnorm. soc. Psychol.*, 1953, **48**, 585–592.
24. TSAO, F. General solution of the analysis of variance and covariance in the case of unequal or disproportionate numbers of observations in the subclasses. *Psychometrika*, 1946, **11**, 107–128.

Received March 4, 1957. Prior publication.

CRITIQUE AND NOTES

ON THE PHENOMENA OF "OPPOSITE SPEECH"[1]

BERNARD KAPLAN

Clark University

IN A RECENT case study, Laffal *et al.* (16) reported a schizophrenic patient exemplifying "an unusual language syndrome" which the authors refer to as "opposite speech." This syndrome "consists basically in the use of 'yes' by the patient when he means 'no,' and vice versa" (16, p. 409). On a level closer to the data, the syndrome seems to be manifested in the interchange of words of contrasting signification; thus, the patient appeared to interchange "right" and "wrong," "do" and "don't," "like" and "hate," "always" and "never," "any" and "none," etc. Such opposite speech phenomena were manifested in both reception and expression, in both oral and written discourse, and in both spontaneous utterances and responses to questions and test problems. The patient was seemingly unaware of the interchanges; furthermore, there was no analogy to this type of speech activity on the level of concrete, practical action.

In their discussion, Laffal and his collaborators emphasized the different needs of the patient which the opposite speech phenomena might have served; they also considered the inadequacies of contiguity-reinforcement theory in accounting for these phenomena. However, with respect to the nature of the syndrome in the light of cognitive development, the authors limited themselves to a few preliminary remarks and did not take up the issue subsequently. Confronted by what Arieti has described as the schizophrenic's "teleological regression" (2, p. 192), they dealt predominantly with the ostensible teleology and secondarily with the nature of the regression.

It is the purpose of this paper to consider "opposite speech" within the categories of a theory of the development of cognition—a theory concerned with general, formal properties of cognitive activity, obtaining in phylogenesis, ontogenesis, cultural evolution, psychopathology, etc. (24, 25, 26, 27).

The fundamental principle of this comparative-developmental approach to cognition is that wherever development occurs, it proceeds from a relatively global and undifferentiated state to one of increasing differentiation, articulation, and integration; on the other hand, where development is undone, there is a return to a greater autonomy of local functions and structures. These correlative principles have the status of heuristic laws (cf. 22); though not themselves subject to empirical test, they are valuable in the formulation of empirically testable generalizations.

One such empirical generalization is that adult schizophrenics, in deviating from normal adult modes of organizing experience, move in the direction of a relative dedifferentiation of their cognitively constructed world, and toward a greater autonomy of local functional patterns organized on a lower level of integration (e.g., toward a greater autonomy of concrete, practical activity with respect to linguistically articulated, conceptual thought).[2]

Insofar as cognitive activity in schizophrenics undergoes this dedifferentiation and disintegration, it will show similarities in structure (cf. 23, p. 27 ff.) to cognitive activity occurring under a variety of conditions and among certain groups of nonschizophrenic individuals. Thus, schizophrenic cognitive activity will tend to show formal similarities to the activities of normal man (*a*) under transitory conditions of stress, anxiety, or distraction; (*b*) in the process of expressing experiences in primitive media of symbolization, e.g., dreams, gestures, etc.; (*c*) in earlier stages of sociocultural linguistic codifications of experience. Schizophrenic cognitive activity may also be expected to show similar formal properties to those reflected in the linguistic articulation of experiences among (*d*) very young children; and (*e*) such organically dedifferentiated individuals as certain aphasics.

There follows (*a*) a critical discussion of the findings presented by Laffal *et al.*; (*b*) an application of the developmental principle to show why "opposite speech" phenomena are regarded as involving a more or less primitive cognitive activity; (*c*) a summary of the empirical evidence

[1] This paper derives, in part, from research and study carried out within a project on normal and pathological symbolization and communication. This project, directed by Heinz Werner and the author, is supported by a grant from The Foundations' Fund for Research in Psychiatry.

[2] Dedifferentiation or regression are not here considered as processes of delamination. Regression may be manifested in a return to certain more primitive levels of cognitive activity without entailing the impairment of "schemata" or "apparatuses" structured in the organism in the course of mastering other, genetically more advanced, operations. Furthermore, regression does not here signify a re-evoking of old content, but rather a return to more primitive forms of thought and patterns of operations (cf. 2, 5).

from each of the five above-mentioned domains in order to evaluate the thesis that "opposite speech" is more characteristic of relatively primitive cognition than of logico-scientific cognition. Finally, (*d*) some of the findings obtained in two recent experimental studies on symbolization will be discussed in order further to support the main thesis of this paper.

From the examples cited by Laffal as illustrations of "opposite speech," there is at least the possibility that these investigators have subsumed under a common rubic essentially dissimilar phenomena with respect to cognitive processes. Thus, when they state that the patient remarked, "I'm hungry," and then refused to take food from his mother, there is at least the possibility that his refusal was based on some delusion with respect to food coming from his mother (cf. 15, p. 40); the refusal does not necessarily warrant the interpretation that the patient really meant "I'm not hungry." Again, when a male schizophrenic, to the statement, *I am a man* responds with "no" (written) and "Nope, I'm no man" (oral), it does not follow that he is confusing or interchanging "no" with "yes"; he may well be deluded with respect to his sex. There are several other examples whose classification involves the same arbitrariness of interpretation, but the above may suffice to ground the criticism that several diversely mediated phenomena may well be grouped by Laffal *et al.* under a single heading.

Leaving aside these questionable instances, there were a number of illustrations provided by Laffal which appear clearly to exemplify "opposite speech" in the sense of involving either (*a*) a lack of differentiation in the use of affirmative and negative forms of judgment, e.g., "I know," I don't know," etc., or (*b*) an undifferentiatedness in linguistic expression for logically opposite meanings, e.g., the use of "like" for both "like" and "hate," etc.

Thus, it seems that "opposite speech" was operative when the patient elaborated on "limping around" by adding "able to walk" or on "disgusted" by remarking "when you're feeling good, happy and contented." Again, it seems that "opposite speech" was in play when the patient completed the sentence *What makes me angry...* with "I feel good, I get very nervous inside, I start to like."

The underlying confusions between affirmation and negation or between opposing predicates seem especially clear in those instances when the patient appeared to be struggling to clarify his thought-expression; for example, when he completed the sentence, *I feel like cursing when...* with "things go good, or if I have any hard luck, get in trouble in any way. Well, if things go right... If I have some plans and things come up to break it... don't come up to break it."[3]

The grounds on which developmental theorists expect opposite speech phenomena to be more characteristic of primitive thought than of relatively advanced, logico-scientific thought may now be briefly considered. As has been mentioned, thought is defined as more primitive the more undifferentiated and unintegrated it is. Opposite speech is presumed to arise because the content to which the linguistic signs refer is in such a global, undifferentiated state that the linguistic vehicles (seemingly discrete) really share the same global referent (3, pp. 182 ff.; 17, p. 18; 19, p. 40; 20, pp. 119 ff.). Concretely, a linguistic sign like "hate" and a sign like "love" both refer to a sphere of "strong affective response" as distinguished from "neutral reaction"; signs like "hot" and "cold" both refer to a semantic sphere of "extreme deviations from normal temperature," etc. It is due to the undifferentiated character of the referent, then, that two signs or linguistic sign complexes, which in mature thought function to articulate and differentiate a global semantic sphere, in primitive thought are used interchangeably. It follows from this characterization that when there is complete and systematic exchange of symbols ("love" used always for "hate" and vice versa), opposite speech is not necessarily manifest. Opposite speech is in play only where the same sign or sign-complex refers now correctly, now incorrectly.

Turning now to the five domains mentioned above, one may examine the evidence to determine whether there is warrant in concluding that such verbal interchanges or undifferentiated symbolizations are more characteristic of these domains than of logico-scientific domain.

First, even in normal adults conditions of stress or anxiety—presumably leading to a dedifferentiation of thought—promote the emergence of "opposite speech" phenomena, e.g., "left" is said for "right," "on" for "off," etc. Many phenomena subsumed by psychoanalysts under the concept of "psychopathology of everyday life" are clearly cases of opposite speech. In this connection Freud remarks, "Numerous observations have taught

[3] It may be noted that in most of the examples provided by Laffal *et al.*, the manifestations of "opposite speech" may have come about through the omission by the patient of "not" or "don't"; in other words, what often appears to be an interchange of opposites may rather be a positive judgment, indifferently signifying both an affirmative and a negative judgment. There are some instances, however, which seem unambiguously to involve an interchange of opposites, e.g., the patient completes the sentence, "When people make fun of me...," with "I start to like."

me . . . that we frequently interchange contrasting words; they are already associated in our speech consciousness; they lie very close together and are easily incorrectly evoked" (9, p. 43).

Second, opposite speech phenomena are characteristically exhibited in the relation between manifest and latent dream contents. Freud, summarizing his observations of dreams, remarks: "The way in which dreams treat the category of contraries and contradictories is highly remarkable. It is simply disregarded. 'No' seems not to exist so far as dreams are concerned. They show a particular preference for combining contraries in a unity, or for representing them as one and the same thing. Dreams feel themselves at liberty, moreover, to represent any element by its wishful contrary; so that there is no way of deciding at a first glance whether any element that admits of a contrary is present in the dream thoughts as a positive or as a negative" (10, p. 318). If one accepts Freud's inquiries as representative and his analyses as correct, it would seem that opposite speech phenomena are characteristic of the dream state.

Third, it has been argued that there is a strong tendency toward opposite speech in early forms of language. Thus Karl Abel (1, pp. 225 ff.) claimed that in the most ancient languages there characteristically appears the use of a single verbal symbol to represent what for Western adult mentality are two contrasting significations, two contraries at the extreme ends of a series of qualities or activities. Even when distinct terms for the two contraries emerge, their kinship is reflected in the fact that the differentiation is achieved by slightly modifying the "mother form" to express one of the contraries, while the original form is retained for the other. Abel illustrates his thesis with reference to the Coptic language: In Coptic, *kef* is used to express both "to take" or "to let lie"; *ken* designates "strong" and/or "weak." The symbols for "stand" and "move" are *men* and *menmen*, respectively—this last example illustrating the primal emergence and representation of more precise determinations of the common, global sphere.[4]

Research on some contemporary "primitive" (cf. 27) languages yield the presence of similar phenomena. For example, Westermann, in his study of the Ewe language (cited in 6) observes that one and the same adverb serves to designate both "yesterday" and "tomorrow"; again, in the Shambala language, the same word refers both to the earliest time and the distant future.[5]

Fourth, in their renowned work, *Die Kindersprache*, Clara and William Stern (21) cite many examples of the use of a single symbol among children for what to normal adults are contrasting significations. For example, their daughter, Hilde at 12½ months used "papa" to designate both her father and mother; at 16 months she used "mama" to designate both her mother and father; at 18 months she interchanged "ja" and "nein" (pp. 29); at 21 months she confused "winter" and "sonne," and also "kalt" and "heiss" (pp. 35 ff.); at two years of age, she would use the adverb of place, "zu," for "zu" and "auf" (pp. 56).

According to the Sterns, the use of antithetical words belongs almost exclusively to the first stages of child language. They differentiate a small number of groups of words which seem to be interchanged with some regularity among the children discussed in the literature. One group, for example, consists of place adverbs of motion. Among German children, the words "an" and "auf" are especially interchanged. A second group involves expressions of activity. For instance, one investigator's daughter used "bi" both for being dressed and being undressed. Another investigator observed that his son at 23 months used the word "moga" for "I can" and "I can't," and also "znam" for "I know" and "I don't know." A third group in which antithetical meanings are expressed by a single verbal symbol comprises adjectives, very frequently adjectives of temperature; the Sterns mention that seven of the children discussed in the literature interchanged "heiss" and "kalt," etc.[6]

Decroly's (8, p. 234) investigations supplement the findings cited by the Sterns and others. In studying his child, E., Decroly found many instances of "opposite speech." Thus, *feme* was used to express both "fermer" (close) and "ouvrir" (open); *cho* signified both "chaud" (warm) and "froid" (cold); *lume* expressed both "allumer" (light up) and "eteindre" (extinguish); *pen* was

[4] According to Abel, the same tendency to express contrasting significations by a single symbol characterizes early stages of the Semitic and Indo-European languages. For a more recent linguistic discussion of such undifferentiated, "neutral" symbols, see (13).

[5] Cassirer, in discussing this lack of differentiation between expressions for past and future, observes that it derives, in part, from the primitive intuition of time as a *thing*. A sharp distinction is made early between the contents of the now and the not-now, but the not-now is not given any further determination (6, p. 220).

[6] The Chamberlains (7), not cited by the Sterns, give some examples, analogous to those quoted by Laffal *et al.*, with respect to their patient's interchange of "any" for "no" or "none." At 31 months, their girl often used "any," "anything," or "anybody" in the sense of "none," "nothing," "nobody"; thus, she remarked, "Any more pickas in zat book" for "No more pictures . . . ," etc.

used for "prendre" (take) and "donner" (give), etc. All of these findings serve to support the thesis that opposite speech is relatively more characteristic of the linguistic activity of younger children than of older children and adults.

Fifth, students of the language behavior of aphasics have often remarked the occurrence of phenomena subsumable under the concept of "opposite speech." One of the earliest investigators, Bastian, cites several cases in which aphasic patients could say both "yes" and "no," but could not always use these terms appropriately (4, pp. 97, 99). Additional examples of such linguistic interchanges may be found in the works of Head (14), Goldstein (12), and Weisenburg and McBride (23).

Head remarks of one of his patients that "he could not grasp the exact meaning of 'to' and 'past' (the hour), 'high' and 'low,' 'up,' and 'down,' 'back' and 'front,' 'right' and 'left' when contrasted with one another. Asked, 'Would you like to stop in or go out?' he would reply, 'Either,' in order to hide his want of complete comprehension" (14, 1, p. 180).

One of the patients examined by Weisenburg and McBride answered "yes, yes" to all questions when he meant "no" (p. 315). On a test for the employment of opposites, this patient, according to the authors, showed "an increasing tendency to lose the controlling idea 'opposite of.'" Another patient tended to use "only the words 'no' and 'yes' and to say 'no' very decidedly sometimes when she obviously meant 'yes'" (23, p. 557). Goldstein, whose conceptual framework is closely related to that presented here, not only observed such linguistic interchanges in his aphasic patients, but pointed out its analogy to child linguistic behavior (cf. 12, p. 39) as has been attempted here with schizophrenics. Assuming that these examples of "opposite speech" in aphasics are more representative of that group than of normals, this provides another source of evidence for assigning "opposite speech" to a relatively lower stratum in the development of linguistic thought organization.

Two experimental studies on symbolization and thought which are relevant to the problem of "opposite speech" may now be discussed. In one part of the first study, carried out by R. Pollack (18), three groups of children (ranging in age from six years, seven months to 12 years, six months) and a group of college students were presented with the task of selecting colors for the representation of 20 mood terms (e.g., happy, sad; nice, nasty, etc.) so chosen that the total list consisted of 10 pairs of opposites, all subsumable under two broad classes of positively toned and negatively toned words. In general, it was found that younger children tended more than older ones to employ the same color to represent what, to adults, are more or less antithetical meanings. Moreover, when the opposition between terms was highlighted by juxtaposing the antonyms, the older children and adults seemed to benefit from the experimental polarization of opposites by more consistently selecting different colors for the antithetical words; the younger children showed no noticeable improvement. The groups were then asked to symbolize the mood terms expressively by drawing linear symbols. Only the adults expressively distinguished positive and negative mood terms by linear schematization when the mood terms were presented in random order. When presented with a juxtaposition of opposites, the youngest group of children alone showed no improvement in the tendency to differentiate the positive and negative opposites. This study points to the relative undifferentiatedness among younger children of what later are given determination as opposites.

The second experimental study, carried out by A. Goldman (11), was concerned with symbolic reception and expression among schizophrenics contrasted with normal adults. In the first part of the study, the two groups of subjects were randomly presented with 20 mood terms to be represented in line patterns; again the total group of words consisted of 10 pairs of opposites. In the second part of the experiment, the subjects were presented with line symbols constructed by the experimenter and asked to match the symbols with the mood terms. Goldman found that schizophrenics, both receptively and expressively, were more likely than normals to utilize the same symbol to express antithetical meanings or sharply contrasting symbols to express the same meanings. Thus, a schizophrenic was significantly more likely than a normal to take on one occasion a curved, light, upward line, and on a second occasion an angular, dark, downward line as expressive of "joy" or to use the light, curved, upward line on one occasion as expressive of "joy" and on a second occasion as expressive of "sadness."

The findings of the experiments by Pollack and Goldman, and the general tenor of the evidence from the five domains in which primitive thought expression might plausibly be thought to obtain, suggest that the processes underlying "opposite speech" occur not only in a few schizophrenics employing a strange means to avoid anxiety or to express hostility, but may be found in any individual, characteristically or momentarily operating under conditions conducive to a primitivization of the level of symbolic articulation and organization of experience. It is hoped that this discussion shows the value of a comparative-developmental analysis of the formal aspects of cognitive functioning as a

complement to dynamic considerations for a fuller understanding of schizophrenic symptomatology (cf. 2, Part III).

REFERENCES

1. ABEL, K. *Linguistic essays*. Boston: Houghton Mifflin, 1882.
2. ARIETI, S. *Interpretation of schizophrenia*. New York: Brunner, 1955.
3. BALDWIN, J. M. *Thought and things*. Vol. 1. New York: Macmillan, 1906.
4. BASTIAN, H. C. *Aphasia and other speech defects*. London, 1898.
5. BAUDOUIN, C. La regression et les phenomènes de recul en psychologie. *J. de Psychol.* 1928, **25**, 795–823.
6. CASSIRER, E. *The philosophy of symbolic forms*, Vol. I (trans. by R. Manheim). New Haven: Yale Univer. Press, 1953.
7. CHAMBERLAIN, A. F., & CHAMBERLAIN, I. C. Studies of a child. *Ped. Sem.*, 1904, **11**, 249–263.
8. DECROLY, O. *Comment l'enfant arrive à parler*, Vol. II. Belgium: Cahiers de la Centrale, 1934.
9. FREUD, S. *Psychopathology of everyday life*. New York: Mentor, 1952.
10. FREUD, S. *The interpretation of dreams* (trans. by J. Strachey *et al.*). London: Hogarth Press, 1953.
11. GOLDMAN, A. E. An investigation of symbolic representation in schizophrenia. Unpublished doctoral dissertation, Clark Univer., 1955.
12. GOLDSTEIN, K. *Language and language disturbances*. New York: Grune & Stratton, 1948.
13. HATZFELD, H. *Leitfaden der vergleichenden Bedeutungslehre*. Hueber: Munich, 1928.
14. HEAD, H. *Aphasia and kindred disorders of speech*. 2 vols. Cambridge: Cambridge Univer. Press, 1926.
15. HILL, L. B. *Psychotherapeutic intervention in schizophrenia*. Chicago: Univer. of Chicago Press, 1955.
16. LAFFAL, J., LENKOSKI, L., & AMEEN, L. "Opposite Speech" in a schizophrenic patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 409–413.
17. MESSER, A. Experimentelle psychologische Untersuchungen uber das denken. *Archiv. f. d. ges. Psychol.*, 1906, **8**, 1–224.
18. POLLACK, R. H. A genetic study of intuitive word meanings. Unpublished doctoral dissertation, Clark Univer., 1953.
19. RIBOT, T. *Essay on the creative imagination* (trans. by A. baron). Chicago: Open Court, 1906.
20. SIGWART, C. *Logic*. Vol. I (trans. by H. Dendy). New York: Macmillan, 1895.
21. STERN, CLARA, & STERN, W. *Die Kindersprache*. Leipzig: Barth, 1907.
22. TOULMIN, S. *The philosophy of science*. London: Hutchinson, 1953.
23. WEISENBURG, T., & MCBRIDE, K. *Aphasia*. New York: The Commonwealth Fund, 1935.
24. WERNER, H. *Comparative psychology of mental development*. (3rd ed.) New York: International Universities Press, 1957.
25. WERNER, H. Microgenesis and aphasia. *J. abnorm. soc. Psychol.*, 1956, **52**, 347–353.
26. WERNER, H. The concept of development from a comparative and organismic point of view. In D. Harris (Ed.), *The concept of development*. Minneapolis: Univer. of Minnesota Press, 1957.
27. WERNER, H., & KAPLAN, B. The developmental approach to cognition: its relevance to the psychological interpretation of anthropological and ethnolinguistic data. *Amer. Anthropol.*, 1956, **58**, 866–880.
28. WHYTE, L. L. *Accent on form*. New York: Harper, 1954.

Received January 28, 1957.

ETHNOCENTRISM AND STIMULUS GENERALIZATION: A REPLICATION AND FURTHER STUDY

FRANKLYN N. ARNHOFF

Mental Health Research Unit, Dept. of Mental Hygiene, Syracuse, N. Y.

In a previous paper (2), the writer reported an exploratory study in which ethnocentrism test scores and error scores on a visual-spatial task of stimulus generalization (SG) were found to be positively related ($r = .44$, $N = 25$) tending to give support to the hypothesis that ethnocentrism represents an "overgeneralization phenomenon." During the course of an extensive investigation of SG, the previous study was replicated with refinements and additional controls. In addition to the visual-spatial task (lights), a more cognitive task (blocks) was added to the present study and both SG tasks were related to the subjects' (*Ss*') scores on the California Ethnocentrism Scale (E scale) (1). As this study was part of a larger study reported elsewhere (3), and the two generalization tasks have been previously described in detail (2, 3, 4, 5), the reader is referred to other sources for descriptions of the tasks and the experimental design.

The volunteer *Ss* were 60 female student nurses, aged 18 through 22. As far as can be determined, the sample was comparable to the one in the previous study (2).

RESULTS AND DISCUSSION

Following administration of the two SG tasks, *Ss* were given the E scale and the CVS Intelligence Scale. Prior to the study proper, all *Ss* had been administered the group form of the Minnesota Multiphasic Personality Inventory (MMPI).

Generalization was defined as the number of errors made on the two tasks. Total error scores were computed for each *S* on each task. These were then correlated with *Ss'* E scale scores, yielding *r*'s of .10 (lights) and .15 (blocks), values falling far short of significance.

IQ was not found to be a significant factor in SG error scores, although its relation to the E scale ($r = -.31$) was consistent with previously reported findings (1).

Test-retest scores on the E scale were found reliable ($r = .85$) and overdefensive responding as measured by the MMPI *K* scale was only minimally related to E scale scores ($r = -.29$), ruling out these factors as explanation for the present failure.

The previous positive findings must therefore probably be attributed to chance. While present results do not preclude the possibility that ethnocentrism is related to overgeneralization in thinking, the functions involved in performance on the tasks investigated do not appear to be on the same psychological continuum.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ARNHOFF, F. N. Ethnocentrism and stimulus generalization. *J. abnorm. soc. Psychol.*, 1956, **53,** 138–139.
3. ARNHOFF, F. N., & LOY, D. L. The relationship between two measures of stimulus generalization: The influence of intelligence upon performance. *Psychol. Rep.*, in press.
4. BROWN, J. S., BILODEAU, E. A., & BARON, M. P. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41,** 52–61.
5. BUSS, A. H. Stimulus generalization as a function of clinical anxiety and direction of generalization. *J. abnorm. soc. Psychol.*, 1955, **50,** 271–273.

Received November 8, 1956.

SOCIAL DESIRABILITY AND PROBABILITY OF ENDORSEMENT OF ITEMS IN THE INTERPERSONAL CHECK LIST[1]

ALLEN L. EDWARDS

University of Washington

One of the most widely used methods for obtaining self-descriptions of personality is the personality inventory. As is well known, the typical personality inventory consists of a number of statements or items to which a person is asked to respond by indicating all those that he believes characterize himself. His responses to the statements thus provide his self-description. If a large number of individuals are given the same inventory, it is possible to find the proportion endorsing each statement, that is, the proportion who say that a given statement is applicable in self-description. Each of these proportions may be regarded as the probability that a given statement will be answered "Yes," "Agree," or "True," depending on the form of the inventory.

It has been found that it is possible to predict these probabilities of endorsement with a rather high degree of accuracy. The predictions are possible because the probabilities of endorsement are linearly related to another independently determined characteristic of the statements, namely, their social desirability scale values. The relationship between these two variables has been found consistently for a variety of statements investigated.

The social desirability scale value of a personality statement refers to its location on a psychological continuum of social desirability. A judging group, for example, is asked to make comparative judgments of the degree of social desirability of each of the statements in a set. On the basis of these judgments, and by means of one of the psychological scaling methods, a scale value is obtained for each statement on a psychological continuum of social desirability. High scale values on this continuum represent statements that are judged as socially desirable and low scale values represent statements that are judged as socially undesirable.

Suppose we scale the statements in a given

[1] This paper was presented at the American Psychological Association meetings, Chicago, Illinois, September 1, 1956. The study was supported by a grant from the Graduate School of the University of Washington.

personality inventory for social desirability in terms of the judgments of one group of subjects. The inventory is then administered to a new and independent group, that is, one other than that used in determining the social desirability scale values, with the usual instructions to provide self-descriptions. For this second group, the proportion endorsing each of the statements in the inventory is obtained. It is then possible to determine the relationship between probability of endorsement and social desirability scale value for the set of statements.

Following the procedure just described, Edwards (1) found a product-moment correlation of .87 between probability of endorsement and social desirability scale value for a set of 140 personality statements. This finding has since been substantiated by Kenny (4) and by Hanley (3) and, indirectly, by Rosen (10). Edwards' research was done with statements subsequently used in developing the PPS (2), a forced-choice inventory in which statements are paired on the basis of their social desirability scale values. Kenny's research was done with a set of personality statements originally used in a study by Zimmer (12). The statements used by Hanley were selected from three of the MMPI scales. Rosen's study was also done with statements from the MMPI.

In all of the studies mentioned above, the subjects given the personality inventories were asked to sign their names on their answer sheets. That this requirement may be an important factor in the relationship between probability of endorsement and social desirability scale value is indicated by Meehl and Hathaway (8). After surveying research by Maller (7), Olson (9), and Spencer (11) on self-descriptions made under anonymous conditions, Meehl and Hathaway conclude that requiring subjects to sign their names to their self-descriptions has been shown to have a definite influence on the descriptions obtained. The present study was undertaken, therefore, to determine whether the relationship between probability of endorsement and social desirability scale value would be found when subjects made their assessments anonymously.

METHOD

The statements used in the study were those appearing in the 128-item Interpersonal Check List (ICL). The ICL was developed by staff members of the Kaiser Foundation Research Project in Psychology. It has been described in an article by LaForge and Suczek (5) and has been prepared in booklet form by Leary (6). The ICL can be used for purposes of both self-assessment and assessment by others. In self-assessments, the *S* checks each item he regards as self-descriptive. The ICL thus closely resembles the typical personality inventory in which the *S* responds "Yes" or "No" to the individual items in the inventory. The ICL statements were selected for investigation because it seemed of value to extend our knowledge of the relationship between probability of endorsement and social desirability scale value with respect to new and different statements which ha not been used in any of the previous studies.

The 128 items in the ICL were given to a group of 58 male and 39 female college students who were asked to make comparative judgments of the social desirability of each of the items. Judgments were made in terms of a nine-interval scale, the first interval being defined as highly socially undesirable and the last as highly socially desirable, following the same instructions used in an earlier study (1). Using these judgments, a scale value for each item can be obtained by the method of equal-appearing intervals.

Before actually calculating scale values of the items, the distributions of judgments for the males and females were examined to determine whether the judgments obtained from the two groups could be pooled. This was done by determining the interval in which the scale value of each statement would fall for each group of judges considered separately. The correlation between these scale intervals over the 128 statements was .90, a value sufficiently high to warrant pooling the two distributions. On the basis of the judgments of the combined group of 94 to 97 judges, equal-appearing interval scale values were found for the 128 items. The scale value for a particular item corresponds to the median of the distribution of judgments for the item on the nine-interval scale. The number of judges varies for the items because of failure to obtain complete sets of judgments for all 58 male judges.

The ICL was then given to students registered in speech classes at the University of Washington with standard instructions for obtaining self-descriptions, except that they were told *not* to put their names on their test booklets. In addition, the students were told that their instructor was not interested in identifying anyone's answers and that the data were to be used in a study of sex differences. It was hoped that these instructions would offer some assurance to the students that their answers would indeed remain anonymous.

RESULTS

Complete records were obtained from 67 female and 66 male students. For each of these two groups the frequency of endorsement of each item was found and transformed to a proportion. These proportions were regarded as the probabilities of endorsement for the various items. The product-moment correlation between the female and male probabilities over the set of 128 items was .95. Because of the high relationship between the two sets of probabilities, the data were pooled and a new set of probabilities obtained for the combined group of 133 *S*s.

Probability of endorsement of an item was then plotted against the previously determined social desirability scale value of the item. The plot was linear and the product-moment correlation between the two variables was .83.

DISCUSSION AND CONCLUSIONS

If anonymity has any influence on self-descriptions, it would seem that the influence should be such as to produce more frank evaluations. Under anonymous conditions, therefore, one might expect responses to personality statements to be less related to the social desirability scale values of the statements than when *Ss* are asked to sign their names to their self-assessments. The correlation of .83 obtained between the two variables of interest in the present study is, however, not significantly different from the value of .87 reported previously for a different set of statements and a different sample of *Ss* and where the assessments were not made anonymously.

The correlation reported, of course, does not bear upon the problem of whether or not we would have obtained a significantly higher correlation for the same set of statements if *Ss* had been asked to sign their names to their self-descriptions. More important would seem to be the fact that the correlation obtained in the present study is of sufficient magnitude to indicate that assurance of anonymity does not eliminate nor drastically change the nature of the relationship previously found between probability of endorsement and social desirability scale value.

REFERENCES

1. EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
2. EDWARDS, A. L. *Manual for the Personal Preference Schedule.* New York: Psychological Corp., 1954.
3. HANLEY, C. Social desirability and responses to items from three MMPI scales: *D*, *Sc*, and *K*. *J. appl. Psychol.* 1956, **40**, 324–328.
4. KENNY, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. *J. consult. Psychol.*, 1956, **20**, 315–318.
5. LAFORGE, R., & SUCZEK, R. F. The interpersonal dimension of personality: III. An interpersonal check list. *J. Pers.*, 1955, **24**, 94–112.
6. LEARY, T. The Interpersonal Check List. Berkeley, Calif. Psychol. Consult. Service. (Undated)
7. MALLER, J. B. The effect of signing one's name. *Sch. and Soc.*, 1930, **31**, 882–884.
8. MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525–564.
9. OLSON, W. C. The waiver of signature in personal reports. *J. appl. Psychol.*, 1936, **20**, 442–450.
10. ROSEN, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, **52**, 151–158.
11. SPENCER, D. Frankness of subjects on personality measures. *J. educ. Psychol.*, 1938, **29**, 26–35.
12. ZIMMER, H. Self-acceptance and its relation to conflict. *J. consult. Psychol.*, 1954, **18**, 447–449.

Received Janaury 14, 1957.

SOCIAL SETTING AND CONFORMITY TO A LEGAL REQUIREMENT[1,2]

ABRAM M. BARCH, DON TRUMBO, AND JOHN NANGLE

Michigan State University

THE PURPOSE of the study was to note the effect on a subject's behavior of observing others conform or fail to conform to a legal requirement. The specific hypothesis under test was that the signalling behavior of the driver of a motor vehicle preparing to turn at an intersection would be related positively to the signalling behavior of the driver of an immediately preceding motor vehicle.

The research was stimulated by the work of Blake and his associates concerned with the influence of social background or social setting on conformity behavior (1, 2, 3, 4), but the data are also relevant to other formulations of imitation and social conformity. In the situation studied, there is a minimum of interaction between the subject and the model, and the subjects can be assumed to have had considerable opportunity to learn what the typical behavior of other individuals in situations similar to the test situation would be. Unlike the settings investigated by Blake *et al.*, conformity or nonconformity in the present situation is to a legal requirement.

[1] The study was supported by the Highway Traffic Safety Center of Michigan State University.

[2] The authors are indebted to Dr. Charles Hanley for statistical advice and assistance.

METHOD

The Observations

Preliminary observations indicated that the variation in turn-signalling present in the naturalistic situation was sufficient for the purpose of the study.

Four different intersections in the Greater Lansing area were observed for a total of 61 hours

over a four-week period in the summer of 1956. The particular hours during which the intersections were observed depended on considerations involved in a broader study of turn-signalling behavior; all data were collected during daylight hours and in good weather (no fog or rain). Intersections studied included two instances of two two-lane roads meeting at a two-way stop, one four-lane highway intersecting a two-lane road at a stop signal, and one six-lane divided highway intersecting a three-lane road at a stop signal.

During any period of observation, the observer noted all cars moving from a given direction on a given leg of the intersection and recorded the following information for all turning cars: (*a*) presence or absence of a signal to turn; (*b*) direction of turn; (*c*) presence of another car 100 feet or less behind the turning car when it began its turn; (*d*) sex of the driver. (At two of the seven legs observed, only cars turning in a given direction were recorded because of the low occurrence of turns in one direction and/or the difficulty of accurately noting turns in both directions.)

The observer stationed himself near the intersection where he could best make the observations required. In most cases he was diagonally across the intersection from the observed traffic movement. At other times he stood about 100 feet ahead of the intersection along the lane(s) being observed. Data collected under control conditions designed to eliminate the possibility that drivers might note the presence of observers indicated that the visible presence of observers did not affect the percentage of cars signalling for a turn. In a reliability check, agreement between two observers for the same cars was 99 per cent or better for all categories except "the presence of a following car" (94 per cent).

During the periods of observation, 4,229 male drivers and 1,004 female drivers of passenger cars turned at the locations under study.

Criteria for Scoring Conformity

Michigan state law requires that intention to turn be signalled by either hand signal or electric signal but does not specify the distance this signal must be given prior to turning or make a distinction between a signal for a left turn and one for a right turn. Therefore, a driver was designated as signalling if he blinked his left or right turn signal light or gave any hand signal, except a hand signal for stopping, regardless of the direction of turn.

RESULTS

The first step in the analysis consisted of extracting the appropriate pairs from the field observation data. The criteria for selection were: (*a*) both cars turned at the intersection under study; (*b*) the following car was 100 feet or less behind the lead car when the lead car began its turn; and (*c*) no cars were in the same lane between the two turning cars. A given car could appear in only one pair of cars; i.e., it could not be a following car for one pair and the lead car for the next pair. The total number of such pairs was 1,195. Of these, 723 pairs had a conforming model and 472 pairs had a nonconforming model.

TABLE 1
REACTIONS OF SUBJECTS TO SIGNALLING (S) AND NONSIGNALLING (NS) MODELS

Behavior of Model		Behavior of Subject							
		High Conformity Legs				Low Conformity Legs			
		Right Turn		Left Turn		Right Turn		Left Turn	
		S	NS	S	NS	S	NS	S	NS
S	*n*	140	74	167	62	48	100	77	55
	%	65	35	73	27	32	68	58	42
NS	*n*	84	64	36	21	42	103	64	58
	%	57	43	63	37	29	71	52	48

The signalling behavior of the follower was significantly related at the .01 level to the signalling behavior of the lead car ($\chi^2 = 15.78$, 1 *df*). This positive relationship, although highly significant, was found to be a weak one. The strength of the association as estimated by tetrachoric *r* was .19.

A future analysis of the data was felt desirable since the general level of conformity (percentage of all turning cars signalling without regard to presence of a model car) varied with the direction of turn and the leg of the intersection. Women also tended to signal more frequently than men. However, the difference was small and the pattern of relationships within the data for women was quite similar to that for men. Therefore, men and women drivers were combined in all comparisons.

Table 1 presents the number and percentage of subjects signalling or failing to signal in relation to the signalling behavior of the model under various conditions of direction of turn of the subject and the general conformity level characteristic of the intersection leg at which the turn was observed. General conformity level for an intersection leg was taken as the percentage of turn-signalling for all cars turning in a given direction without regard to the presence of a model car. Intersection legs were classified as high conformity if the percentage of turning-signalling for left turns was between 62.5 and 88 and the percentage of signalling for right turns was between 37.5 and 62.5. Low conformity legs had a signalling percentage between 37.5 and 62.5 for left turns and one between 20 and 37.5 for right turns.

Table 2 presents the results of the χ^2 analyses

TABLE 2
STATISTICAL ANALYSES OF THE INFLUENCE OF THE MODEL

Respondent Categories	χ^{2a}	p	r_{tet}
All cases	15.78	.01	.19
All left turns	6.62	.05	.20
All right turns	4.82	.05	.14
All cases at high conformity legs	6.75	.05	.18
All cases at low conformity legs	1.17	—	.09
Left turns at high conformity legs	1.67	—	.18
Right turns at high conformity legs	2.43	—	.14
Left turns at low conformity legs	0.66	—	.09
Right turns at low conformity legs	0.27	—	.07

[a] All analyses have 1 *df*.

of the influence of the signalling behavior of the model on the subject's signalling for various combinations of conditions. Tetrachoric r is provided for each combination to show the strength of the association. In all comparisons the relationship was positive to the hypothesis of influence. The strength of the relationship was low for all analyses, ranging from .07 to .20.

An attempt was made to increase the possible identification of the subject with the model by considering only those pairs in which both members of the pair were of the same sex and turned in the same direction. There were 691 such pairs. The χ^2 for this analysis, as would be expected, was significant at the .01 level ($\chi^2 = 7.33$, 1 *df*). However, the tetrachoric r was found to be .17, essentially the same as that obtained for the other comparisons.

The results supported the hypothesis that the behavior exhibited by others in a situation involving a legal norm influences a subject's behavior. The weakness of the demonstrated influence should be considered in light of the lack of formal relationships or of interaction, except in a minimal sense, between the model and the subject and the subject's ample opportunity to observe the behavior of others in the same or similar situation prior to the test observation.

SUMMARY

Observations made under normal traffic conditions were analyzed to determine the influence on the signalling behavior of automobile drivers of the signalling behavior of other drivers. A weak but significant positive relationship was found between observation of conformity of others to a legal requirement and self conformity.

REFERENCES

1. BLAKE, R. R., & MOUTON, JANE S. Present and future implications of social psychology for law and lawyers. *Symposium Issue, Emory Univer. J. Public Law*, 1955, **3**, 352–369.
2. FREED, A. M., CHANDLER, P. J., MOUTON, JANE S., & BLAKE, R. R. Stimulus and background factors in sign violation. *J. Pers.*, 1955, **23**, 499.
3. LEFKOWITZ, M., BLAKE, R. R., & MOUTON, JANE S. Status factors in pedestrian violations of traffic signals. *J. abnorm. soc. Psychol.*, 1955, **51**, 704–706.
4. ROSENBAUM, M., & BLAKE, R. R. Volunteering as a function of field structure. *J. abnorm. soc. Psychol.*, 1955, **50**, 193–196.

Received January 16, 1957.

PUBLIC AND PRIVATE CONFORMITY UNDER DIFFERENT CONDITIONS OF ACCEPTANCE IN THE GROUP[1]

HERBERT MENZEL
Columbia University

DITTES and Kelley (2) recently reported an experiment on conformity to group norms which showed, among other things, that the subject's experimentally induced feeling of acceptance within the group had a rather different effect on his private and public expressions of conformity to the group's previously established judgment. It was found, in brief, that individuals whose feeling of being accepted in the group was high felt the greatest freedom to express any disagreement with the group's judgment publicly. Individuals whose feeling of acceptance was low exhibited high public conformity but showed little conformity in their private expressions. Individuals who were made to feel an "average" level of acceptance exhibited a "high degree of genuine adherence to the norms . . . extending even to conditions of privacy." The

[1] This paper may be identified as Publication No. A-236 of the Bureau of Applied Social Research, Columbia University, and is an outgrowth of a study conducted by James Coleman, Elihu Katz, and the present writer.

present note addresses itself exclusively to the similarity or difference between public and private conformity, and not to the determination of private conformity as such.

DATA

The conclusions of Dittes and Kelley in this respect can be applied in slightly modified form to data, originally secured for another purpose (see 1, 4), on the use of new drugs by physicians. Eighty-five per cent of the general practitioners, internists, and pediatricians practicing in four selected cities were interviewed regarding their use of a certain widely used family of drugs, comprising several new variants which had made their appearance successively in the course of recent years. At the same time, the records of prescriptions written by these same doctors were examined in the local pharmacies for sampling dates extending over a period of 15 months. One may think of the physician's writing of prescriptions as his private behavior, and of his declarations in the course of an interview as the analogue of "public" expressions in the sense of the Dittes and Kelley experiment. So defined, "public" and "private" data are available on three items of behavior: (*a*) the date on which the doctor used "gammanym," the newest variant of the family of drugs, for the first time; (*b*) which of the three chief available variants he favored at the time the interviews were conducted; (*c*) which of the two chief variants then available he favored 15 months earlier.

RESULTS AND DISCUSSION

Existence of Norms

In the fluid situation created by the recurrent appearance of new drugs, there is no established norm on record as to which of two or three competing variants a physician should prefer, nor as to the "right" date on which to introduce a new drug into one's practice. What the norm in these respects was at the time the interviews were conducted can, however, be inferred from the prevailing direction of the discrepancies between private performance and public accounts of behavior: (*a*) One-half of the physicians stated as the date on which they first used the newest drug a date earlier—often many months earlier—than that established by the prescription search (Table 1).[2] (*b*) Two-thirds of the physicians whose prescription record at the time of interviewing showed predominant use of the oldest drug stated that they preferred one of

TABLE 1
DATE OF INTRODUCTION OF A NEW DRUG: INTERVAL BETWEEN INTERVIEW STATEMENTS AND PRESCRIPTION RECORD

Date According to Interview	Percentage of Doctors (N = 70)
Follows prescription date by 2 months or more	24%
Falls within one month's distance of prescription date	27
Precedes prescription date by 2–7 months	25
Precedes prescription date by more than 7 months	24
	100%

TABLE 2
DRUG VARIANT FAVORED AT TIME OF INTERVIEW

Variant Favored According to Interview Statement	Variant Favored According to Prescription Record		
	Oldest	Middle	Newest
Oldest	33%	8%	6%
Middle	25	32	18
Newest	42	60	76
Total	100%	100%	100%
N	(24)	(25)	(61)

[2] The extent of this discrepancy can hardly be accounted for by the sampling of days in the prescription search. (Three consecutive days per month were sampled.) It is true that the presumed starting date of gammanym use, obtained by recording prescriptions on only a sample of days, can err only by being too late, and not by being too early. It is, however, possible to estimate the size of that error, since not only prescriptions for the new drug were recorded, but also prescriptions for its predecessors. It was found that during the average two-month period the proportion of doctors who were missed by the prescription search—i.e., for whom prescriptions for neither the new nor the older drugs were recorded on the sampling days falling within that period—was 23.2% (with a standard deviation of .06%). This may be taken as an estimate of the proportion of doctors who were missed by the prescription search during the two months starting with their respective introduction of the new drug. But since few doctors, if any, made a complete switch from older drugs to newer drugs, one must add to the estimated 23.2% who were altogether missed by the prescription search during these two months the doctors who were covered by the prescription search, but happened to prescribe only older drugs on the sampling days in question. This is estimated at 4.6% of all the doctors in the sample. Thus, if all the doctors had stated their introduction date accurately during the interview, the proportion who would, due to the sampling of days, nevertheless appear to have up-dated themselves by two months or more may be estimated at 23.2% + 4.6% or 27.8%. Actually, as Table 1 shows, 49% appear to have up-dated themselves to this extent.

The above estimate of 4.6% is based on the following considerations. Of all the prescriptions written for this family of drugs by doctors during the sample days within two months after their respective first *sampled* prescription of gammanym, 67% were for older drugs. The proportion of their prescriptions which ordered the older drugs during the two months following their first

TABLE 3
DISCREPANCIES IN DATE OF FIRST PRESCRIPTION OF A NEW DRUG, BY SOCIOMETRIC STATUS

Introduction Date According to Interview Statement	Number of Designations Received as Advisor		
	None	One or Two	Three or More
Earlier than by prescription record	63%	33%	36%
Same as by prescription record[a]	26	33	21
Later than by prescription record	11	33	43
Total	100%	100%	100%
N	(35)	(21)	(14)

[a] Allowing a one-month tolerance.

TABLE 4
DISCREPANCIES IN DATE OF FIRST PRESCRIPTION OF A NEW DRUG, IN RELATION TO SOCIOMETRIC STATUS, WITH ACTUAL PERFORMANCE HELD CONSTANT

Percentage of Doctors who Stated a Drug Introduction Date Earlier (Later) Than the Median Date Stated by Those who Actually Introduced the Drug Simultaneously with Them	Number of Designations received as Advisor		
	None	One or Two	Three or More
Earlier than median	60%	45%	36%
Later than median	40	55	64
Total	100%	100%	100%
N	(35)	(21)	(14)

the newer drugs. Sixty per cent of those whose current prescription record showed predominant use of the middle one of the three drugs stated that they preferred the newest (Table 2). (*c*) A substantial portion (29%) of the 52 doctors who, according to their prescription record, had prescribed predominantly the oldest variant 15 months before the interviews took place, failed to make any mention of this drug when recounting the history of their use of drugs of this general kind.

From these results it may be concluded that there prevailed, at the time of the interviewing, norms in these communities of physicians which placed high value on (*a*) having introduced the newest drug early; (*b*) currently favoring the newest drug over its predecessors; (*c*) having favored the middle drug over the oldest 15 months earlier (i.e., before the appearance of the newest drug). Thus, with respect to all three items of behavior, the preponderant direction of the discrepancies between the prescription record and the interview statements is such that the interview statements make the doctor appear more "up-to-date." This finding supports the inference that the prevailing norm in these medical communities at the time of interviewing was generally favorable to "modernity," at least in the use of drugs. The findings to be reported below, however, are of interest even if the above three sets of discrepancies merely express three separate norms relating specifically to the use of the particular drugs in question.

Conformity and Acceptance in the Group

When the physician's standing in his medical community is taken into account, it is found that the discrepancies in the prevailing direction occur most frequently among the low-status physicians, who tend to make themselves appear more up-to-date than they really are. An individual's standing in the community of his colleagues is measured by the number of designations accorded him in response to the sociometric question: "When you need information or advice about questions of therapy, . . . on whom are you most likely to call?" It is chiefly those seldom or never named in answer to this question whose interview statements make them appear more up-to-date than their prescription record.

Table 3 shows this relationship for the date on which the doctor used the newest of the drugs for the first time. The contrast is quite marked, but is partly due to a ceiling effect: high-status individuals did in fact tend to be more up-to-date in the timing of their first use of this new drug, and hence had less opportunity to deviate, in their interview statements, in the up-to-date direction. However, the phenomenon persists even when actual performance is held constant, as shown in Table 4. Doctors were first divided into two-month intervals according to their drug introduction date as established by the prescription search. Within each such two-month interval, the median interview statement as to the introduction date of the drug was then determined. Each doctor was then classified according to whether his own interview statement gave a date later or earlier than the median statement of doctors who actually introduced the drug within the same two-month inter-

actual prescription of gammanym was probably not far different. The average doctor wrote 7.8 prescriptions for this family of drugs on the sampling days falling within two months of his first sampled prescription of gammanym, and probably as many on the sampling days falling within two months of his first actual prescription of gammanym. The probability of all 7.8 of these prescriptions having been for the older drugs is therefore estimated as the seventh power of .67, which is .06. Applying this factor to the estimated proportion of doctors not altogether missed by the prescription search during the two months starting with their introduction of gammanym (100% − 23.2% = 76.8%) yields a figure of 4.6%.

TABLE 5

DISCREPANCIES IN DRUG FAVORED AT THE TIME OF INTERVIEWING, IN RELATION TO SOCIOMETRIC STATUS

	Number of Designations Received as Advisor:		
	None	One or Two	Three or More
Percentage who stated they favored a *newer* variant than prescription record shows, among those who prescribed predominantly:			
the oldest variant	67% (12)	77% (9)	33% (3)
the middle variant	64% (11)	63% (8)	50% (6)
Percentage who stated they favored an *older* variant than prescription record shows, among those who prescribed predominantly:			
the middle variant	9% (11)	0% (8)	17% (6)
the newest variant	21% (24)	29% (14)	23% (13)

TABLE 6

DISCREPANCIES IN DRUG FAVORED 15 MONTHS PRIOR TO TIME OF INTERVIEWING, IN RELATION TO SOCIOMETRIC STATUS

	Number of Designations Received as Advisor		
	None	One or Two	Three or More
Percentage who stated they favored the newer variant, among those who prescribed predominantly the oldest	33% (24)	30% (20)	13% (8)

val as he did. It is seen that even when possible ceiling effects are eliminated by holding the date of introduction according to the prescription record constant, those of low standing in the group show a quite disproportionate tendency to report an introduction date which precedes that established from the prescription record.[3]

Table 5 records the results obtained with respect to the second item of behavior: variant favored at the time the interviews were conducted. The necessity to control for actual performance here makes for very small base figures, and the results are somewhat less consistent than in Table 4, but they are substantially of the same order. Those of high standing in the group consistently "up-date" their behavior less than do those of low standing, although those of intermediate standing do not always fall into the expected place. The two last rows of Table 5 show that this difference is specific to distortions in the direction of the norm; the relatively few distortions that occur in the opposite direction are, if anything, more prevalent among those of high standing.

Table 6 records the results for the third item of behavior: variant favored, among the two then available, 15 months prior to the time of interviewing. Once again, those of high standing in the community of colleagues only seldom reported having used, 15 months before the interviews took place, a newer drug than predominated among their recorded prescriptions, while those of low standing frequently made such reports.

DISCUSSION

Thus, with respect to all three items of behavior on which information is available, doctors of low standing in the community frequently reported, during an interview, more up-to-date behavior

[3] None of the main results (Tables 4, 5, and 6) are statistically significant in the usual sense, but generalizability of the descriptive results beyond the four cities surveyed could not be claimed in any case. (For Table 4, collapsed to 2 x 2 format, chi square [corrected for continuity] = 1.721; $df = 1$, $p = .096$ for a one-tailed test. [See 5, Table 7, p. 125; 6, pp. 108–110].) If the results were based on a sample randomly drawn from a defined and much larger population, observations on the whole population might very well fail to confirm them. In fact, however, the observations were made on 85% of the target population (general practitioners, pediatricians, and internists practicing in four cities). There is therefore little doubt that the findings would be confirmed by observations on the entire target population. On the other hand, neither the sample nor the target population can be claimed to represent any larger population, regardless of the outcome of significance tests.

It is quite another matter that the results are, of course, intended as contributory evidence of the more general applicability of Dittes' and Kelley's conclusions. Establishment of such a proposition ultimately depends on confirmation in a variety of concrete situations (with diverse populations, norms, indicators of acceptance, and expressions of conformity), rather than on the generalizability of one particular finding to a larger population. They are here shown to apply to three items of behavior in one particular situation rather different from the original experiment. The position taken here is essentially that expressed at some length in Lipset *et al.* (3, pp. 427–432).

than is indicated by the prescription record, while doctors of high standing seldom did so. Since, as was shown above, the general norm favored being "up-to-date" in the three matters under investigation (and perhaps in other maters as well), these results seem to be close parallels to the corresponding findings of the Dittes and Kelley experiment.

One may note that the high-status doctors not infrequently deviated in their "public" (i.e., interview) statements from their prescription record in the "wrong" direction, that is, so as to appear *less* up-to-date than they actually were. It is not clear from Dittes' and Kelley's report whether individuals who agreed with the group privately ever expressed disagreement in public. If not, the contrast may be due to the fact that in the present study the norm of the group was not "crystallized and registered," while the doctors' statements were subject to a certain amount of unmotivated error due to faulty recall, which played no part in Dittes' and Kelley's experiment. It is also possible that there existed, uniquely among the high-status individuals, a competing norm in favor of conservatism; but this seems an unlikely explanation in view of the simultaneous finding that high-status individuals tended to be more up-to-date in fact than low-status individuals, at least in the date of their introduction of the newest of the drugs. (Why high-status individuals were more up-to-date in their actual behavior will be discussed elsewhere. See also [1].)

SUMMARY

An experiment by Dittes and Kelley (2) had shown that individuals who felt accepted in their group felt relatively free to express any disagreement with the group's judgment publicly, while those whose feeling of acceptance was low exhibited much higher conformity in their public than in their private expressions. Data from a survey among physicians provide an analogue in the form of interview statements about prescriptions of new drugs (public expressions) and the doctor's actual prescription record (private behavior). A norm of being up-to-date with respect to three choice situations is inferred from the prevailing direction of discrepancies between private performance and public accounts. Acceptance in the group was measured sociometrically. In all three choice situations, those of low acceptance reported more up-to-date behavior than indicated by the prescription record, while doctors of high acceptance seldom did so.

REFERENCES

1. COLEMAN, J. S., KATZ, E., & MENZEL, H. The diffusion of an innovation among physicians. *Sociometry*, in press.
2. DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100–107.
3. LIPSET, S. M., TROW, M. A., & COLEMAN, J. S. *Union democracy*. Glencoe, Ill.: Free Press, 1956.
4. MENZEL, H., & KATZ, E. Social relations and innovation in the medical profession: the epidemiology of a new drug. *Publ. Opin. Quart.*, 1956, **19**, 337–52.
5. PEARSON, E. S., & HARTLEY, H. O. *Biometrika tables for statisticians*. Cambridge: Univer. Press, 1954.
6. SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Received February 1, 1957.

RESOLUTION OF APPROACH-APPROACH AND AVOIDANCE-AVOIDANCE CONFLICTS[1]

ABE ARKOFF

University of Hawaii

CONSISTENT with the deductions of Lewin (5), Hovland and Sears (4) found that avoidance-avoidance conflicts produced more blocking and compromise responses in the manual motor situations they studied than did approach-approach ones. Barker (1), however, found no consistent evidence that choice time and "vicarious trial and error behavior" were greater in children choosing between paired repellent liquids than between attractive ones. In a second experiment (2), he presented college students with paired hypothetical alternatives involving personal characteristics and environmental conditions and found that the frequency of choices marked "uncertain" was greater for negative than positive alternatives. The present investigation attempted to devise comparable approach-approach and avoidance-avoidance situations that would involve the subjects emotionally and permit clear-cut measures of conflict behavior.

[1] The author wishes to acknowledge the generous assistance of Dr. John M. Digman.

METHOD

Forty male and 40 female students from the introductory psychology course served as subjects (*Ss*). Conflicts were created by systematically pairing with one another seven personal characteristics: adjustment, attractiveness, health, intelligence, popularity, talent, and wealth. Two types of pairings were employed, one in which the *S* was required to designate the alternative that he would rather have in greater degree than he had at present (approach-approach conflict), and a second in which he was required to designate the alternative that he would rather have in lesser degree than he had at present (avoidance-avoidance conflict). In the 42 conflicts presented, 21 described approach-approach situations and 21 described avoidance-avoidance ones. Each choice was typed on a separate 3 x 5 card with the question "Which would you rather be?" centered above, and the alternatives boxed below on left and right portions of the card. An example of an approach-approach conflict follows (for the avoidance-avoidance analogue "less" was substituted for "more"):

Which would you rather be?

More attractive than you are now.	More intelligent than you are now.

In the first part of the experiment, the experimenter (*E*) arranged the 42 conflict cards in random order and placed them face down in a single stack before the *S* who was given written instructions as follow:

> Each card describes a conflict which you must resolve. When the experimenter gives the signal, turn up the topmost card and study the conflict presented. If your choice is the alternative on the left, place the card, face down, on the table to the left of the pile. If your choice is the alternative on the right, place the card, face down, on the table to the right of the pile.
>
> Pay no attention to the timing. Take as much or as little time as you like with each card.
>
> Imagine that each conflict *really* confronts you. Be sure that your choice is the one that you would *really* make if you *really* had to decide.

As soon as the *S* turned up the conflict card (and before he actually began to read the conflict) the *E* began the timing. Timing was stopped as soon as the card was face down and flat on the table.

The conflict cards were then placed once more in random order, and the *S*, without timing, was given written instructions as follow:

> Divide the cards into two piles of 21 cards each. In one pile place the cards describing the conflicts which you found relatively easier to resolve. In the other pile place the cards describing the conflicts which you found relatively harder to resolve. You may proceed to do this in any fashion you wish.

RESULTS

Two measures of conflict behavior were available: the amount of time taken to resolve the two types of conflict, and the number of conflicts of each type judged to be easier to resolve.

The mean resolution times taken for each type of conflict by the male and female groups are contained in Table 1. The results of an analysis of variance, presented in Table 2, indicate that the difference between the time taken to resolve the approach-approach and avoidance-avoidance conflicts was significant at the .01 level of confidence. The interaction between sex and the experimental condition was not significant. The data were further analyzed for possible differences in variability between sexes; at the .05 level of confidence, the hypothesis of equal variability could not be rejected (F = 1.03 for approach-approach scores and 1.48 for avoidance-avoidance scores, df = 39 and 39).

The mean numbers of each type of conflict judged easier to resolve by the male and female groups are contained in Table 3. An analysis of the difference in variability between sexes showed that at the .01 level of confidence the males were more variable than the females (F = 1.98, df = 39 and

TABLE 1

MEAN RESOLUTION TIMES (IN SEC.) FOR APPROACH-APPROACH AND AVOIDANCE-AVOIDANCE CONFLICTS OF MALE AND FEMALE SUBJECTS

Sex	Conflict	
	Approach-Approach	Avoidance-Avoidance
Male	128.92	164.32
Female	115.08	142.00
Both	122.00	153.16

TABLE 2

ANALYSIS OF VARIANCE OF RESOLUTION TIME FOR AVOIDANCE-AVOIDANCE AND APPROACH-APPROACH CONFLICTS

Source of variation	*df*	Mean square	*F*
Between conflicts	1	38844.06	110.28*
Between sexes	1	13086.30	3.14
Conflict × sex	1	718.26	2.04
Between *Ss* within groups	78	4172.52	
Conflict × *Ss* within groups	78	352.22	
Total	159		

* Significant at the .01 level.

TABLE 3

MEAN NUMBER OF APPROACH-APPROACH AND AVOIDANCE-AVOIDANCE CONFLICTS JUDGED EASIER TO RESOLVE

Sex	Conflict	
	Approach-Approach	Avoidance-Avoidance
Male	12.72	8.28
Female	12.55	8.45
Both	12.64	8.36

39). A t test of the type suggested by Gronow (3) for use with groups differing in variability indicated that at the .05 level of confidence the null hypothesis in respect to sex differences could not be rejected ($t = .20$, $df = 39$ and 39). Assuming the null hypothesis tenable, the two sex groups were then combined to test the hypothesis that the mean number of approach-approach conflicts judged easier to resolve minus the mean number of avoidance-avoidance conflicts similarly judged was zero; the results indicated that the null hypothesis could be refuted considerably beyond the .01 level of confidence ($t = 5.74$, $df = 79$).

CONCLUSION

The results of the present experiment are in agreement with Lewin's deduction that approach-approach conflicts are more easily resolved than are avoidance-avoidance conflicts. Approach-approach conflicts required significantly less time to resolve than did avoidance-avoidance conflicts, and significantly more approach-approach conflicts were judged easier to resolve than were avoidance-avoidance conflicts. Males and females did not appear to differ in the conflict behavior except that in judging the difficulty of conflicts, males proved to be significantly more variable than females.

REFERENCES

1. BARKER, R. G. An experimental study of the resolution of conflict by children. In Q. McNemar & Maude A. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942. Pp. 13–34.
2. BARKER, R. G. An experimental study of the relationship between certainty of choice and the relative valence of the alternatives. *J. Pers.*, 1946, **15**, 41–52.
3. GRONOW, D. G. C. Test for the significance of the differences between means in two normal populations having unequal variances. *Biometrika*, 1951, **38**, 252–256.
4. HOVLAND, C. I., & SEARS, R. R. Experiments on motor conflict. I. Types of conflict and their modes of resolution. *J. exp. Psychol.*, 1938, **23**, 477–493.
5. LEWIN, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.

Received October 15, 1956.

PERSONALITY CHANGE AS A RESULT OF AN EXPERIMENTAL CHANGE OF ENVIRONMENTAL CONDITIONS

ARNOLD S. TANNENBAUM

Survey Research Center, University of Michigan

THE QUESTION of the consistency of adult personality is a continuing problem in the field of personality theory. Kelly, in his longitudinal study of marital adjustment, found evidence of a remarkable degree of consistency as well as some evidence of significant changes over a 20-year period (2). One of the questions that remains is the extent to which personality change, when it occurs, can be understood in terms of the effect of specific environmental conditions. The present note is intended to report results bearing on this question from an experimental study in a large clerical organization. Other aspects of this study have been reported earlier (4, 5, 7).

Two experimental groups were created that differed in their distribution of control. In one, called the autonomous program, the amount of control exercised by lower-level employees was increased and that of the upper hierarchical levels decreased. In the other, called the hierarchical program, the amount of control exercised by the rank and file was decreased and that of the upper levels increased. The program for each group of approximately 200 female employees continued for over a year. A paper-and-pencil questionnaire designed specifically to measure 26 personality trends to which each of the programs had some degree of relevance was administered near the beginning of the experimental period, and about a year later at the end. Results reported earlier (7) indicated that subjects' reactions to the work programs can be understood partly in terms of the relationships between their personality structures and the nature of the program in which they were located. It is of further interest, however, to investigate personality as a dependent variable; the extent to which subjects' personality trends are found to change in predictable ways as a result of their experiences in the experimental programs.

It was initially assumed that personality trends are quite stable and would be unlikely to change as a result of experimental treatment.[1] Nevertheless, the two programs were seen to have profound effects on the subjects' attitudes toward the company and satisfactions in it. In addition, the duration of the experiment, as well as the seriousness with which the subjects reacted to it, raised the question of its possible effects on the relevant

[1] For the conception of personality used, see Allport, F. H. (1).

TABLE 1

BEFORE-AFTER CORRELATIONS AND INCIDENCE OF CHANGE

(Change in the predicted direction significant at the .05 level is indicated by +; change in a direction opposite to that predicted by −)

	Before-after r (N = 195)[a]	Autonomous Program (N = 115)[a]	Hierarchical Program (N = 80)[a]
Autonomous Trends			
1. assume responsibility (anticipating, preparing for, and accepting the consequences of one's activities)	.59	n.s.	n.s.
2. be creative (develop new patterns; to use one's imagination; to think up new and original ways of doing things)	.56	n.s.	n.s.
3. increase the variety (number of types) of satisfactions in any given situation (to enjoy as many different things as possible in the same situation, e.g., to "mix business with pleasure")	.41	n.s.	n.s.
4. understand and explain the reasons for one's behavior and opinions (to understand and explain one's feelings)	.48	n.s.	n.s.
5. help others	.59	n.s.	n.s.
6. understand others' viewpoints and feelings (to know why people feel or act as they do; to understand the reasons for others' behavior)	.60	n.s.	+
7. be independent of control figures (to be free of direction and control of persons in authority)	.54	n.s.	+
8. express one's considered judgments (to think through things carefully and state one's opinions)	.78	n.s.	+
9. base one's actions on one's own critical judgments and evaluations (to carefully evaluate and decide things for oneself)	.38	−	+
10. act on the same level as others (to treat and be treated by others as an equal)	.52	n.s.	n.s.
11. understand relations between and reasons for things (to understand the basis for things; to figure things out; to know "why")	.72	−	+
12. do one's own thinking (not be suggestible; to decide things for oneself; to draw one's own conclusions)	.40	n.s.	+
13. take the initiative with others (to start things; to organize things)	.42	n.s.	−
Hierarchical Trends			
14. avoid emotional involvements (to avoid the expression and reception of emotional acceptance; to keep emotionally distant from others; to avoid being "warm" in relationships with others; to avoid affection)	.65	n.s.	n.s.
15. conform to the wishes of control figures (to please persons in authority)	.60	n.s.	n.s.
16. adhere rigidly to specific rules or directions (to follow rules to the letter; to be given directions for doing things)	.58	n.s.	n.s.
17. act inferior to control figures (to be humble in relation to persons in authority)	.60	+	n.s.
18. avoid committing oneself (to keep one's opinions to oneself; to keep from getting oneself "out on a limb"; to keep people from knowing where one stands on controversial questions)	.40	n.s.	n.s.
19. keep one's roles (behavior categories, relationships) at any given time distinct from other roles (to keep various parts of one's life separated)	.40	n.s.	+
20. admire or respect control figures or symbols (to look up to persons in, or symbols of, authority)	.68	n.s.	n.s.
21. be superior to others (to be better than others; to be more outstanding than others)	.69	+	−
22. be efficient (not waste time or effort)	.58	+	−
23. be submissive to control figures (to submit to dominant leaders; to obey strong and forceful authority figures)	.45	n.s.	+
24. show self-discipline (to show self control; to keep one's behavior disciplined)	.42	+	−
25. depend on others (to receive direction from others)	.48	n.s.	n.s.
26. obey rules and follow directions	.58	n.s.	n.s.

[a] In some cases the N's are slightly less than this value due to a very small proportion of NA's—never more than 4%.

personality variables. A number of hypotheses regarding change in personality trends were therefore formulated on the basis of the assumption that environmental conditions may increase or decrease the characteristic energic (homeostatic) level of trends on the basis of a simple reward-and-punishment principle: characteristics that are given opportunities for expression tend to increase in potency; trends that are given only minimal opportunities for expression tend to decrease.

The 26 trends measured by the personality questionnaire were divided into two groups of thirteen: "autonomous trends," which were likely to be given opportunities for expression in the autonomous program (and not in the hierarchical), and "hierarchical trends," which were likely to be

given opportunities for expression in the hierarchical program (and not in the autonomous). It was hypothesized that (*a*) in the autonomous program, autonomous trends should tend to increase in potency and hierarchical trends tend to decrease; (*b*) in the hierarchical program, hierarchical trends should tend to increase in potency and autonomous trends tend to decrease.

Care was taken in framing the questions used to measure the personality trends to make them as general as possible and distinct from the company situation. Responses to these questions are therefore not likely to reflect the subjects' changing attitudes toward the experimental programs or toward the company. Five questions were employed to measure each trend. For example, one of the questions for the trend, "trying to be independent of control figures," was "How much do you believe that people generally require someone to tell them what to do?" For the trend, "trying to understand others' viewpoints and feelings," we asked, "How much does it bother you when you do not understand the viewpoints and feelings of people you associate with?" For the trend, "trying to keep one's role at any given time distinct from other roles," we asked, "Generally, how good an idea do you think it is to choose very close friends from among persons you work with?" For the trend, "trying to do one's own thinking," we asked "How hard do you find it to disagree with others even in your own thinking?" For the trend, "trying to act inferior to control figures," we asked, "Do you think that people who have reached high office in government have a right to feel that they are better than most people?" For the trend, "trying to be submissive to control figures," we asked, "Do you think that for a truly happy marriage that in major issues the husband should tell the wife what to do?"

RESULTS

Table 1 shows the before-after correlations (test-retest reliabilities) for each of the trends, and the direction of changes significant at the .05 level or less. A plus sign indicates that the change is in the direction predicted; a minus sign, a change in the opposite direction. Although all of the predictions are directional, two-tailed rather than one-tailed tests are employed to permit ascribing levels of confidence to changes contrary to hypothesis.[2]

Twelve changes in the predicted direction prove significant at the .05 level of confidence. Six changes, significant at the .05 level occur in a direction opposite to that predicted.[3] The *combined* results cannot be assessed statistically since the trend measures are not independent, but it seems reasonable to conclude that some change has occurred and that it is predominantly in the predicted direction.

The preponderance of changes in the direction of decreasing trend potency is hard to explain. Most of the "correct" changes involve the *decrease* of autonomous trends in the hierarchical program or the *decrease* of hierarchical trends in the autonomous program. Likewise, five of the six "incorrect" changes involve the *decrease* of autonomous trends in the autonomous program or hierarchical trends in the hierarchical program. No single trend is found to change significantly in the predicted direction in *both* programs.

Although somewhat ambiguous, the data seem to indicate that measurable change can be effected by a persisting change in environmental conditions. Furthermore, the change seems partly explicable in terms of the movement of personality toward equilibrium with its environment. A type of organism-environment reciprocity is suggested in which the organism selects, reacts, and adjusts to its environment, so as to optimize personality "closure" while at the same time modifying itself within limits so as to further increase trend expression. This integral character of inner needs and outer conditions, as a general psychological principle is consistent with the position taken by Murphy: "A personality is a structured organism-environment field, each aspect of which stands in dynamic relation to each other aspect. There is organization within the organism and organization within the environment, but it is the cross-organization of the two that is investigated in personality research" (6, p. 8).

REFERENCES

1. ALLPORT, F. H. Teleonomic description in the study of personality. *Charact. Pers.*, 1937, **5**, 202–214.
2. KELLEY, E. L. Consistency of the adult personality. *Amer. Psychologist*, 1955, **10**, 659–681.
3. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
4. MORSE, NANCY, & REIMER, E. The experimental change of a major organizational variable. *J. abnorm. soc. Psychol.*, 1956, **52**, 120–129.
5. MORSE, NANCY, REIMER, E., & TANNENBAUM, A. S. Regulation and control in hierarchical organizations. *J. soc. Issues*, 1951, **7**, (3), 41–48.
6. MURPHY, G. *Personality, a biosocial approach to origins and structure*. New York: Harper, 1947.
7. TANNENBAUM, A. S., & ALLPORT, F. H. Personality structure and group structure: an interpretative study of their relationship through an event structure hypothesis. *J. abnorm. soc. Psychol.*, 1956, **53**, 272–280.

Received February 18, 1957.

[2] The following statistical formula, designed specifically to test the significance of change, was employed: $t = (A - D - 1/\sqrt{A + D})$ (3, p. 207).

[3] The pattern of results remains essentially the same when evaluated in terms of the .01 level of confidence. In this case, eight changes prove "correct" and four "incorrect."

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

GARDNER LINDZEY
University of Minnesota

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: BARBARA CUMMINGS; FRANCES H. CLARK; SADIE J. DOYLE, SARAH WOMACK

VOLUME 55

1957

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.

AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1957 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 55

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

PAGE

ALEXANDER, I. E., & BLACKMAN, S. **Castration, Circumcision, and Anti-Semitism** 143

ALLPORT, G. W., & PETTIGREW, T. F. **Cultural Influence on the Perception of Movement: The Trapezoidal Illusion Among Zulus** 104

ALPER, THELMA G. **Predicting the Direction of Selective Recall: Its Relation to Ego Strength and *N* Achievement** 149

APPLEZWEIG, M. H. See MOELLER, G.

ARKOFF, A. **Resolution of Approach-Approach and Avoidance-Avoidance Conflicts** 402

ARNHOFF, F. N. **Ethnocentrism and Stimulus Generalization: A Replication and Further Study** .. 393

BACON, MARGARET K. See BARRY, H., III.

BARCH, A. M., TRUMBO, DON, & NANGLE, J. **Social Setting and Conformity to a Legal Requirement** .. 396

BARRY, H., III, Bacon, MARGARET K., & CHILD, I. L. **A Cross-Cultural Survey of Some Sex Differences in Socialization** 327

BERKOWITZ, L. **Effects of Perceived Dependence Relationships Upon Conformity to Group Expectations** 350

BERKOWITZ, L. See LUNDY, R. M.

BINDER, A., MCCONNELL, D., & SJOHOLM, NANCY A. **Verbal Conditioning as a Function of Experimenter Characteristics** 309

BINDON, D. MARJORIE. **Make-A-Picture Story (MAPS) Test Findings for Rubella Deaf Children** 38

BLACKMAN, S. See ALEXANDER, I. E.

BLOCK, J. **A Study of Affective Responsiveness in a Lie-Detection Situation** 11

BOREN, J. J. See SIDMAN, M.

BOSSOM, J., & MASLOW, A. H. **Security of Judges as a Factor in Impressions of Warmth in Others** 147

BROWN, R. W. **Linguistic Determinism and the Part of Speech** 1

BRUNER, J. S., & PERLMUTTER, H. V. **Compatriot and Foreigner: A Study of Impression Formation in Three Countries** 253

CARON, A. J., & WALLACH, M. A. **Recall of Interrupted Tasks Under Stress: A Phenomenon of Memory or of Learning?** 372

CHILD, I. L. See BARRY, H., III.

CLEVELAND, S. See WARE, K. E.

COLLIER, R. M. **Consciousness as a Regulatory Field: A Theory of Psychotherapy** 275

CROW, W. J. **The Effect of Training Upon Accuracy and Variability in Interpersonal Perception** ... 355

DAVIS, R. H. **A Further Study of the Effect of Stress on Palmar Prints** 132

DECHARMS, R. **Affiliation Motivation and Productivity in Small Groups** 222

DENNIS, W. **Uses of Common Objects as Indicators of Cultural Orientations** 21

DITTES, J. E. **Galvanic Skin Response as a Measure of Patient's Reaction to Therapist's Permissiveness** 295

DAULNY, D. E., JR. **Avoidance Learning of Perceptual Defense and Vigilance** 333

EDWARDS, A. L. **Social Desirability and Probability of Endorsement of Items in the Interpersonal Check List** 394

ERRATA 273

EXLINE, R. V. **Group Climate as a Factor in the Relevance and Accuracy of Social Perception** 382

FESHBACH, S., & SINGER, R. D. **The Effects of Fear Arousal and Suppression of Fear Upon Social Perception** 283

FISHER, S. See WARE, K. E.

FUNKENSTEIN, D. H. See KING, S. H.

GAGE, N. L., LEAVITT, G. S., & STONE, G. C. **The Psychological Meaning of Acquiescence Set for Authoritarianism** 98

GRICE, G. R. **On the Relations Between Discrimination Reaction Time, Anxiety, and Intelligence** .. 147

GRISELL, J. L. See ROSENBAUM, G.

HARE, A. P. **Situational Differences in Leader Behavior** 132

HARRIS, D. B. **A Scale for Measuring Attitudes of Social Responsibility in Children** 322

HART, I. **Maternal Child-Rearing Practices and Authoritarian Ideology** 232

HARVEY, O. J. See HOVLAND, C. I.

HARWOOD, A. See MEDNICK, S. A.

HILL, W. F. See SCHEIN, E. H.

HORST, P. See STROTHER, C. R.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. **Assimilation and Contrast Effects in Reactions to Communication and Attitude Change** 244

KAMIYA, J. See STONE, P.

KAPLAN, B. **On the Phenomena of "Opposite Speech"** 389

KELLEY, H. H., & LAMB, T. W. **Certainty of Judgment and Resistance to Social Influence** 137

KING, S. H., & FUNKENSTEIN, D. H. **Religious Practice and Cardiovascular Reactions During Stress** 135
LAMB, T. W. See KELLEY, H. H.
LANZETTA, J. T., & ROBY, T. B. **Group Learning and Communication as a Function of Task and Structure "Demands"** 121
LEAVITT, G. S. See GAGE, N. L.
LESSER, G. S. **The Relationship Between Overt and Fantasy Aggression as a Function of Maternal Response to Aggression** 218
LEVENTHAL, D. See MORAN, L. J.
LEVENTHAL, H. **Cognitive Processes and Interpersonal Predictions** 176
LEVIN, H., & TURGEON, VALERIE F. **The Influence of the Mother's Presence on Children's Doll Play Aggression** 304
LIVSON, N., & MUSSEN, P. H. **The Relation of Ego Control to Overt Aggression and Dependency** 66
LUBIN, A. See SCHEIN, E. H.
LUNDY, R. M., & BERKOWITZ, L. **Cognitive Complexity and Assimilative Projection in Attitude Change** 34
LYKKEN, D. T. **A Study of Anxiety in the Sociopathic Personality** 6
MACCOBY, ELEANOR E., & WILSON, W. C. **Identification and Observational Learning from Films** 76
MACHI, V. S. See WILLIAMS, R. J.
MACKAVEY, W. R. See ROSENBAUM, G.
MANDLER, G., & PARNES, ELIZABETH W. **Frequency and Idiosyncracy of Associative Responses** 58
MARTIN, B. See TRUAX, C. B.
MASLOW, A. H. See BOSSOM, J.
MCCONNELL, D. See BINDER, A.
MCDIARMID, C. G. See STEINER, I. D.
MCGAUGHRAN, L. S. See MORAN, L. J.
MEDNICK, MARTHA T. **Mediated Generalization and the Incubation Effect as a Function of Manifest Anxiety** 315
MEDNICK, S. A., HARWOOD, A., & WERTHEIM, J. **Perception of Disturbing and Neutral Words Through the Autokinetic Word Technique** 267
MENZEL, H. **Public and Private Conformity Under Different Conditions of Acceptance in the Group** 398
MEYER, V. **The Treatment of Two Phobic Patients on the Basis of Learning Principles** 261
MILLON, T. **Authoritarianism, Intolerance of Ambiguity, and Rigidity Under Ego- and Task-Involving Conditions** 29
MILTON, G. A. **The Effects of Sex-Role Identification Upon Problem-Solving Skill** 208
MOELLER, G., & APPLEZWEIG, M. H. **A Motivational Factor in Conformity** 114
MORAN, L. J., MCGAUGHRAN, L. S., & LEVENTHAL, D. **Intra-Individual Consistency in Conceptualization** 194
MUSSEN, P. H. See LIVSON, N.
NANGLE, J. See BARCH, A. M.
PARNES, ELIZABETH W. See MANDLER, G.
PERLMUTTER, H. V. See BRUNER, J. S.
PETTIGREW, T. F., See ALLPORT, G. W.
PRENTICE, N. M. **The Influence of Ethnic Groups** 270
REED, M. R. **The Masculinity-Femininity Dimension in Normal and Psychotic Subjects** 289
ROBY, T. B. See LANZETTA, J. T.
ROSENBAUM, G., GRISELL, J. L., & MACKAVEY, W. R. **The Relationship of Age and Privilege Status to Reaction Time Indices of Schizophrenic Motivation** 202
ROSNER, S. **Consistency in Response to Group Pressures** 146
SALZINGER, K. **Shift in Judgment of Weights as a Function of Anchoring Stimuli and Instructions in Early Schizophrenics and Normals** 43
SAMELSON, F. **Conforming Behavior Under Two Conditions of Conflict in the Cognitive Field** 181
SCHAIE, K. W. See STROTHER, C. R.
SCHEIN, E. H., HILL, W. F., WILLIAMS, H. L., & LUBIN, A. **Distinguishing Characteristics of Collaborators and Resisters Among American Prisoners of War** 197
SCOTT, W. A. **Attitude Change Through Reward of Verbal Behavior** 72
SHERIF, M. See HOVLAND, C. I.
SIDMAN, M. & BOREN, J. J. **The Relative Aversiveness of Warning Signal and Shock in an Avoidance Situation** 339
SIEGEL, ALBERTA E., & SIEGEL, S. **Reference Groups, Membership Groups, and Attitude Change** 360
SIEGEL, S. See SIEGEL, ALBERTA E.
SILLER, J. **Socioeconomic Status and Conceptual Thinking** 365
SILVERMAN, R. E. **The Manifest Anxiety Scale as a Measure of Drive** 94
SINGER, R. D. See FESHBACH, S.
SJOHOLM, NANCY A. See BINDER, A.

SMITH, E. E. **The Effects of Clear and Unclear Role Expectations on Group Productivity and Defensiveness** 213
STAATS, A. W. **Learning Theory and "Opposite Speech"** 268
STEINER, I. D., & MCDIARMID, C. G. **Two Kinds of Assumed Similarity Between Opposites** 140
STONE, G. C. See GAGE, N. L.
STONE, PHIL, & KAMIYA, J. **Judgments of Consensus During Group Discussion** 171
STROTHER, C. R., SCHAIE, K. W., & HORST, P. **The Relationship Between Advanced Age and Mental Abilities** 166
TAMKIN, A. S. **Selective Recall in Schizophrenia and Its Relation to Ego Strength** 345
TANNENBAUM, A. S. **Personality Change as a Result of an Experimental Change of Environmental Conditions** 404
TRUAX, C. B. **The Repression Response to Implied Failure as a Function of the Hysteria-Psychasthenia Index** 188
TRAUX, C. B., & MARTIN, B. **The Immediate and Delayed Effect of Failure as a Function of Task Complexity and Personalization of Failure** 16
TRUMBO, DON. See BARCH, A. M.
TURGEON, VALERIE F. See LEVIN, H.
WALLACH, M. A. See CARON, A. J.
WARE, K. E., FISHER, S., & CLEVELAND, S. **Body-Image Boundaries and Adjustment to Poliomyelitis** 88
WERTHEIM, J. See MEDNICK, S. A.
WILLIAMS, H. L. See SCHEIN, E. H.
WILLIAMS, R. J., & MACHI, V. S. **An Analysis of Interperson Correlations Among Thirty Psychotics** 50
WILSON, W. C. See MACCOBY, ELEANOR E.
WORCHEL, P. **Catharsis and the Relief of Hostility** 238
WYLIE, RUTH C. **Cognitive Set and Motivational Factors in the Perception of Neutral and Threat-Related Stimuli** 227

ARE THERE GAPS IN YOUR FILES OF APA JOURNALS

SALE & BACK ISSUES

THEN HEAR THIS . . .

The American Psychological Association announces a sale during the period October 1957 through March 1958. Of the following journals, all *available* issues in the volumes for the years preceding 1951 will be offered at a price of only 50¢ (foreign, 60¢) per issue:

American Psychologist
Journal of Abnormal & Social Psychology
Journal of Applied Psychology
Journal of Comparative & Physiological Psychology (1947–1950 only)
Journal of Consulting Psychology
Journal of Experimental Psychology
Psychological Abstracts
Psychological Bulletin
Psychological Index (a few complete volumes, some shopworn)
Psychological Monographs
Psychological Review

Not *all* issues in *all* volumes are available. But—ORDER NOW before more back issues go out of print. From our available stock we will complete as much of your order as possible at this reduced price and for this limited period.

Delivery: 6 to 8 weeks • No dealer or quantity discounts

After this sale, for the years preceding 1948, journals will be available only on microfilm and microcard.

Order from:

American Psychological Association
Department BB, 1333 Sixteenth Street, N.W.
Washington 6, D. C.

PSYCHIATRY

JOURNAL FOR THE STUDY OF INTERPERSONAL PROCESSES

Contents of No. 3, Vol. 20, August, 1957

ROBERT A. COHEN AND ERVING GOFFMAN: *Introduction—On Some Convergences of Sociology and Psychiatry*

PAUL DANIEL GUIRDON: *Techniques of Sociotherapy*

MAXWELL JONES: *The Treatment of Personality Disorders in a Therapeutic Community*

STEWART E. PERRY AND GERTRUDE N. SHEA: *Social Controls and Psychiatric Theory in a Ward Setting*

MORRIS S. SCHWARTZ: *Patient Demands in a Mental Hospital Context*

KAI T. ERIKSON: *Patient Role and Uncertainty—A Dilemma of the Mentally Ill*

CHARLOTTE GREEN SCHWARTZ: *Perspectives of Deviance—Wives' Definitions of their Husbands' Mental Illness*

HANS H. STRUPP: *A Multidimensional System for Analyzing Psychotherapeutic Techniques*

EDITORIAL NOTES

BRIEF COMMUNICATIONS

BOOK REVIEWS

A clear, explicit description of the healthy personality, as well as a frame of reference for evaluating mental health, is provided by Sidney M. Jourard in his new book PERSONAL ADJUSTMENT: An Approach Through the Study of Healthy Personality. In the author's words, this book represents *"an attempt to describe the 'symptoms' of healthy personality. In addition to description, an attempt is made wherever possible to explain how personality-health is achieved; i.e., to point out the independent variable of which it is a function . . .*

"In most works which deal with abnormal psychology, psychiatry, or mental health, a person is adjudged healthy, or 'normal', if he has no clinical symptoms of pathology. Yet we know now that personality-health is more than mere freedom from symptoms. Just as there is a difference in medicine between freedom from medical symptoms of disease and optimum physical health, so there is a difference between a personality which is 'not sick' and one which is healthy . . ."

Quoted from the preface to
PERSONAL ADJUSTMENT:
An Approach Through the
Study of Healthy Personality

By SIDNEY M. JOURARD
Associate Professor at the University of Florida

To be published Spring, 1958

Examination copies
available upon publication
from The Macmillan Company,
60 Fifth Avenue, New York 11,
New York.